AN ANALYSIS OF STABILITY PROPERTIES IN EARNED VALUE

MANAGEMENT'S COST PERFORMANCE INDEX AND EARNED

SCHEDULE'S SCHEDULE PERFORMANCE INDEX

THESIS

Jacob L. Petter, Second Lieutenant, USAF

AFIT-ENV-14-M-49

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-14-M-49

AN ANALYSIS OF STABILITY PROPERTIES IN EARNED VALUE MANAGEMENT'S
COST PERFORMANCE INDEX AND EARNED SCHEDULE'S SCHEDULE
PERFORMANCE INDEX

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Jacob L. Petter, BS

Second Lieutenant, USAF

March 2014

AFIT-ENV-14-M-49

AN ANALYSIS OF STABILITY PROPERTIES IN EARNED VALUE MANAGEMENT'S COST PERFORMANCE INDEX AND EARNED SCHEDULE'S SCHEDULE PERFORMANCE INDEX

Jacob L. Petter, BS

Second Lieutenant, USAF

Approved:

_____//signed//_____	14 Feb 2014
Lt Col Jonathan D. Ritschel, Ph.D (Chairman)	Date
_____//signed//_____	14 Feb 2014
Dr. Edward D. White, Ph.D (Member)	Date
_____//signed//_____	14 Feb 2014
Michael J. Seibel, Civ., USAF (Member)	Date

Abstract

The concept of a Cost Performance Index (CPI) "stability rule" originated with the seminal article from Christensen and Payne in 1992 and has become routinely cited by subsequent academic literature and EVM authors. A literature review reveals that the definition of what constitutes "stability" has morphed over time, with three separate definitions of "stability" permeating the literature. Additionally, while the original Christensen research found the cumulative CPI to be stable in 86% of DoD contracts (from 155 analyzed) at the 20 percent completion point, more recent research from Henderson and Zwikael (2008) questioned the generalizability of these findings. This research reexamines the question of CPI stability in a modern portfolio of DoD contracts utilizing all three definitions of stability. Next, this research examines potential stability in the Earned Schedule SPI(t) metric. The second stage of this research investigates whether there is a difference in CPI or SPI(t) stabilities between military services, contract types, acquisition life-cycle phases, or platforms. Comparison analysis executes tests on the median stability value for each category of contract. This research finds that CPI stability *both contradicts and supports* the stability rule depending on the stability definition used. The SPI(t) exhibits similar stability traits to CPI stability.

Acknowledgements

Table of Contents

List of Figures

List of Tables

AN ANALYSIS OF STABILITY PROPERTIES IN EARNED VALUE

MANAGEMENT'S COST PERFORMANCE INDEX AND EARNED SCHEDULE'S

SCHEDULE PERFORMANCE INDEX

I. Introduction

Background

In a fiscal environment where each Department of Defense (DoD) program has to fight for funding, key decision makers rely more heavily on measurements that can accurately predict if a program will be successful in adhering to the budget or become a financial catastrophe. If they can conclude early on whether the program will succeed or not, they can make better decisions about funding that program and spend the government's money more wisely.

In the Earned Value Management (EVM) system, the Cost Performance Index (CPI) is an efficiency index used to determine how well a program performs. The CPI is a ratio between the Budgeted Cost of Work Performed (BCWP) and the Actual Cost of Work Performed (ACWP). The BCWP equals the sum of all budgets for completed work packages, or the "earned value" (DCMA, 2006: 90). ACWP represents the cost actually incurred to accomplish the work completed within a specific time frame. CPI, therefore, is the ratio of work performed to the actual costs. It indicates the value of every dollar of work accomplished. A CPI of 0.9 means the program receives ninety cents of budgeted value for every dollar spent, whereas a CPI of 1.1 means the program receives $1.10 of budgeted value for every dollar spent (GAO, 2009: 259). The CPI measures the performance of a program thus far, but can it predict future performance? CPI gives the

efficiency of a program based on the data from the past. If the CPI could provide an estimate of how the program performs in the future, the CPI could help planners see into the future.

Researchers previously examined this possible attribute of CPI. The seminal work originates with Kirk Payne in 1990, where he performed a small scale study with data from twenty-six cost performance reports (CPRs) on seven aircraft in the U.S. Air Force Systems Command Aeronautical Systems Division (Payne, 1990). In a more robust study in 1993, David Christensen and Scott Heise studied the CPI of 155 contracts associated with forty-four different DoD programs. They found that in 86% of the contracts the cumulative CPI (CPI using all available data to date as opposed to data from only that month) stabilized for a program once that program reached the 20% completion point (Christensen and Heise, 1993). They defined stability as having a range of less than 0.2, meaning the minimum and maximum CPI (from the 20% completion point to the end point) had a difference of less than 0.2. These findings became known within the EVM community as the "stability rule." As time went on, the stability rule morphed into a rule of thumb generalized to all programs using EVM even though Christensen and Heise did not claim generalizability in their results (Fleming and Koppelman, 2008: 17).

Fifteen years later, Kym Henderson and Ofer Zwikael re-examined this "stability rule" using a different set of data consisting of forty-five projects dealing with information technology and construction in the United Kingdom, Israel, and Australia (Henderson and Zwikael, 2008). In contrast to Christensen and Heise, they found that the CPI did not stabilize until much later than the 20% completion point. Henderson and Zwikael also attempted an analysis of DoD specific contracts. They did not, however,

have access to primary data from DoD EVM databases to conduct their analysis. Instead, they used secondary data from Michael Popp of U.S. Navy Air Command's (NAVAIR) research on CPI. Through visual inspection of Popp's scatter plots, Henderson and Zwikael explained that, "for the DoD project data used by Popp, the CPI stability was … achieved very late in the project life cycle, often as late as 70-80 percent completion" (Henderson and Zwikael, 2008: 10). Henderson and Zwikael concluded that the CPI stability rule from Christensen and Heise's research is invalid because it is not true for all DoD data. Additionally, Henderson and Zwikael challenge the stability rule's generalizability to any program, DoD or non-DoD. The contradictory conclusions between these two research efforts spawned a heated debate within the EVM community. This disagreement warrants further research into the stability rule. Since the last study with primary DoD data was completed twelve years ago (Christensen and Templin, 2002), it also necessitates a new examination of the stability properties with current data to determine if it exists today.

Along with the CPI, the other efficiency index of EVM is the Schedule Performance Index (SPI). SPI, however, has several well documented drawbacks. The SPI is in units of dollars, which can be difficult to understand when discussing schedule performance. Also, SPI becomes inaccurate because its formula always regresses to equal 1.0 at the end of a contract's life even if the program is behind schedule. Because of this, Walt Lipke states, "at some point it becomes obvious when the SV and SPI indicators have lost their management value" (Lipke, 2003: 3). In response to these deficiencies, the concept of Earned Schedule (ES) was developed with a schedule efficiency index where time is the unit of measurement (SPI(t)). Instead of money as the

3

unit of analysis as with original SPI, SPI(t) is in units of time to be more useful to
practitioners over the entire life of the contract. Additionally, SPI(t) does not display the
same regressive (to 1.0) properties as SPI and thus provides greater accuracy (Crumrine
and Ritschel, 2013). Given SPI(t)'s useful properties, its potential for increased usage in
DoD acquisition programs is vast. Henderson and Zwikael, in their 2008 study
mentioned earlier, examined SPI(t) in thirty-seven non-DoD programs to find that it did
not stabilize (2008: 9). It is unknown whether SPI(t) displays stability properties, similar
to those claimed for CPI in US DoD acquisition programs. There is no previous research
in the area of SPI(t) stability using US DoD datasets or on large scale programs.

Problem Statement

The purpose of this research is to re-examine the existence of the CPI stability
rule with contemporary data to determine the percentage complete point where stability is
achieved. The second major component of this research is to ascertain if SPI(t)
demonstrates similar stability trends to CPI. Thus, we attempt to determine if the
stability rule exists for either CPI or SPI(t) and if so, when in a program's life it occurs.
We also compare different categories of contracts to determine if stability properties vary
for different categories.

The benefits to a stable CPI or SPI(t) are many, to include giving the key decision
makers the ability to assess if a program can recover from poor performance in the
beginning of the program's life. For an over-budget program, the decision makers will be
able to say with confidence that the program's performance will not improve and make
decisions accordingly to efficiently use resources (Christensen and Payne, 1992). A
stable CPI also shows the contractor's management skills. At the beginning of a

4

program, learning occurs and variance in performance may be high. But as the program continues, performance should become stable, meaning that the contractors are learning how to be more efficient and make fewer mistakes. A stable CPI shows government leaders that the contractor manages well. Also, the CPI is based on the budget. The budget is the plan or estimate of the contractor's work, so a stable CPI means the contractor's work is following the designated plan (Payne, 1990: 10). This research attempts to determine the existence of CPI stability that would achieve these benefits, as well as evaluate if SPI(t) has stability properties that could provide these same benefits.

Research Questions

Using data from the Defense Acquisition Executive Summary (DAES) database and the EVM Central Repository, an online database for centralized reporting of DoD programs' EVM statistics, we calculate the CPI throughout the life of each contract available. Comparing the CPIs at each point in the program's life determines if the CPI changes more than a specified amount once the program reaches a certain percentage complete. We test if certain groupings (e.g. by contract type, platform, branch of service) have CPIs that stabilize differently. Then, we repeat this process of determining stability with SPI(t). This research strives to provide answers for the following questions:

1. When in a program's life does the CPI tend to stabilize?
2. When in a program's life does the SPI(t) tend to stabilize?
3. What differences in CPI and SPI(t) stabilities exist between branches of the DoD?
4. What differences in CPI and SPI(t) stabilities exist between contract types?
5. What differences in CPI and SPI(t) stabilities exist between life-cycle phases?

6. What differences in CPI and SPI(t) stabilities exist between different military

platforms?

Methodology

We analyze the CPI and SPI(t) data from the 10 percent complete point to the end

point of a program in increments of 5 percent to determine when the CPI and SPI(t)

stabilize. Stability has three definitions. With the first definition, stability is when the

range from the maximum to minimum CPI or SPI(t) is no greater than 0.2. We determine

the percentage of stable contracts at each percent complete increment of 5 from 10% to

85%. We also calculate a 95% confidence interval at each percent complete increment

on the range at that increment. We next examine stability with a second definition, when

the final CPI is less than +/- 0.1 away from the CPI at a certain percent complete. The

third definition of stability is when the final CPI is less than +/- 10% away from the CPI

at a certain percent complete. The calculation of percentage of stable contracts and

confidence intervals for the analyses of these last two definitions of stability remains the

same as with the first definition of stability.

Scope and Limitations

The scope of the thesis spans all Acquisition Category 1 (ACAT 1) DoD

programs from 1987 to 2012 with available EVM data. Therefore, any available program

must be at 10% complete or less by the start of 1987 and at least 85% complete by the

end of 2012 to be included. Twenty five years of data (1987-2012) encompasses four of

the five major acquisition reforms (the Nunn McCurdy Act, Packard Commission,

Defense Acquisition Workforce Improvement Act, and Federal Acquisition Streamlining

Act) to provide an unbiased picture of DoD programs' performances (Ritschel, 2012:

492). This research contains limitations. The data used includes only ACAT 1 DoD

programs. Therefore, the results cannot be generalized to all programs using EVM.

Also, the internal controls over the EVM data, implementation, and surveillance change

over time, so these changes may affect the data.

Thesis Overview

The second section of this research, Chapter 2, provides a literature review of the

CPI stability rule and research on the SPI(t). Next, Chapter 3 outlines the methodology

of research and how we intend to examine the CPI and SPI(t) possible stability properties

with modern data. Chapter 4 lays out the results of the research and any significant

findings. The last chapter summarizes the previous chapters, states the importance of the

results, and suggests ideas for further research in this area.

II. Literature Review

Overview

This chapter explains the literature and background information regarding the CPI and SPI(t). First, an introduction of program management, the Earned Value Management (EVM) system, and the CPI and SPI(t) is provided. Then, an overview of previous research on the CPI and SPI(t) is presented. The overview highlights the issues this thesis addresses, including the existence of the CPI "stability rule" and the lack of SPI(t) research.

Concepts

Program Management.

Program management is the "application of knowledge, skills, and techniques to execute [programs] effectively and efficiently," (PMI, 2013). According to the DoD directive 5000.01 (guidance for the entire Defense Acquisition System), the Program Manager (PM) is "the designated individual with responsibility for and authority to accomplish program objectives for development, production, and sustainment to meet the user's operational needs" (DoD). The three main responsibilities of the PM, according to DoD 5000.01, are the cost, schedule, and performance reporting of a program. The PM reports these to the Milestone Decision Authority (MDA), who is in charge of approving the program so that it can enter the next phase of acquisition. The MDA is also responsible for reporting the program's performance to Congress and other higher authorities (DoD, 2013: 2). Because of these responsibilities, the PM needs a way to measure the cost, schedule, and performance of the program being managed. Earned

Value Management (EVM) is one of the main tools to track the cost, schedule, and performance of a program.

EVM.

This section introduces EVM, provides a short background of the policies involved, and summarizes the main components of the EVM system.

Background.

According to the *Defense Acquisition Guidebook*, the primary guide for the Acquisition workforce in the DoD, EVM is "a management approach that has evolved from combining both government management requirements and industry best practices to ensure the total integration of cost, schedule, and work scope aspects of the program" (Defense Acquisition University, 2013: 11.3.1). EVM originated from the directive-imposed Cost/Schedule Control Systems Criteria (C/SCSC) that the DoD set as the standard for all programs in 1967 (Fleming and Koppelman, 1998: 19). The C/SCSC was a list of 35 criteria that contractors had to meet when under a contract with the US government. The National Security Industrial Association (NSIA) created an updated version of C/SCSC with 32 criteria in 1995 known as the EVM criteria. The DoD endorsed this criteria in 1996 (Fleming and Koppelman, 1998: 20). EVM's 32 criteria require a defined budget and clear objectives for the program (see Appendix A for a complete list of the 32 criteria).

Since the DoD began using EVM, there have been many changes to regulations and how programs are to utilize EVM. Contract Performance Reports (CPRs), formerly known as Cost Performance Reports, contain the EVM data and overall description of the performance of the contract (DCMA, 2006: 91). Therefore, CPRs are where most the

data on contract performance originates. DoD Directive 5000.02 established the requirements of CPRs in 2003. These requirements were subsequently revised in 2005 (USD-AT&L, 2008: 10). The revision applies only to contracts awarded after its release. Therefore, contracts awarded before 2005 but still ongoing at that time did not have to change in accordance with the revision. The Office of the Under Secretary of Defense (Acquisition, Technology, and Logistics) collects all CPRs in the Central Repository (CR). The CR began collecting CPRs in 2007, and it contains CPRs from before and after this Directive 5000.02 revision (USD-AT&L, 2008: 10). Thus, because the implementation process and reporting rules have changed over the years, the quality of EVM data has potentially been affected. This is an important limitation for EMV studies spanning these time periods.

EVM Measurements.

The primary components of EVM include the Budgeted Cost for Work Performed (BCWP), Actual Cost of Work Performed (ACWP), and Budgeted Cost for Work Scheduled (BCWS). The BCWP, also called the "earned value", is the budgeted cost received for the total work completed. The ACWP is the actual cost that the work incurred. The BCWS, also called the "planned value", is the budgeted value of work scheduled to be completed (DCMA, 2006: 90). Table 1 defines these three initial measurements along with other main EVM measurements that stem from these three.

Table 1. Summary of EVM measurements (DAU, 2013)

EVM measurement	Meaning	Formula
BCWP	The earned value, how much budgeted cost the program has gained thus far	Sum of the budgeted cost of all completed work packages
ACWP	The actual cost of the completed work packages thus far	Sum of actual costs of all completed work packages
BCWS	The planned value, how much budgeted value the program should have gained thus far	Sum of the budgeted cost of all work packages scheduled
Cost Variance (CV)	Difference between planned and actual cost accomplishment	BCWP - ACWP
Schedule Variance (SV)	Difference between planned and actual schedule accomplishment, in dollar amount	BCWP - BCWS
CPI	Cost efficiency of a program	BCWP / ACWP
SPI	Schedule efficiency of a program	BCWP / BCWS
Budget at Complete (BAC)	Planned total cost of program	Sum of all BCWS of program
Budgeted Cost for Work Remaining (BCWR)	The budgeted cost of uncompleted work packages to reach program's completion	BAC – BCWP
Estimate at Complete (EAC)	Forecasted total cost of program	Various formulas
To Complete Performance Index (TCPI)	Projects what the CPI will be for the remainder of the project to meet the BAC	Various formulas
Variance at Completion (VAC)	Estimated cost variance at completion of program	BAC – EAC
Percent Complete	Percentage of the entire program that is complete	BCWPcum / BAC
Total Allocated Budget (TAB)	Total of all contract's work budgets	Sum of all budgets
Contract Base Budget (CBB)	Total budget allotted to the contractor	Sum of the negotiated contract cost and authorized undefined work
Management Reserve (MR)	Amount of the budget allotted for unknown costs or risk management	Determined at start of contract

These measurements in Table 1 are all important components of sound EVM program management. This research, however, focuses on the Cost Performance Index (CPI) and the Schedule Performance Index (SPI).

Efficiency Indices: CPI and SPI (and SPI(t)).

Two primary measurements of EVM that together show the overall performance (cost and schedule) of a program are the two efficiency indices, CPI and SPI. They are

called efficiency indices because they indicate the efficiency of the cost and schedule utilization of a program.

CPI.

As described in Chapter 1, the CPI is the "earned value," or BCWP, divided by the actual cost of work performed, or ACWP (DAU, 2013). CPI indicates the value of work accomplished for every dollar spent. A CPI of 0.98 means the program is receiving 98 cents of value for every dollar spent, and a CPI of 1.05 means the program is receiving a dollar and five cents of value for every dollar spent (GAO, 2009: 259). A CPI of 1.0 means the program is getting one dollar of value for every dollar spent. This explains CPI's designation as an efficiency index. It expresses how efficient the program is spending money while indicating whether the program is on, over, or under budget. The CPI is used to track cost performance throughout a program's life.

The CPI can be calculated with either current or cumulative data. The current CPI uses the BCWP and ACWP of the current period (week, month, quarter, etc.) in its formula, while the cumulative CPI uses the BCWP and ACWP of the entire program's life up to the current date. Both the current and cumulative CPIs have valuable uses. In major defense acquisition programs, the cumulative CPI is involved in calculating the Estimate at Complete (EAC_{CPI}), the estimated final cost of a program, which has been demonstrated as the "reasonable lower bound to the final cost of a defense contract" (Christensen, 1996: 7). Therefore, the CPI helps determine a sound estimate of the entire cost of a contract. The cumulative CPI is the CPI used for this research.

SPI.

The SPI is the other efficiency index. It is the BCWP divided by BCWS (DAU, 2013). It determines the efficiency at which scheduled work is being accomplished. An SPI below 1.0 indicates the program is behind schedule, while an SPI above 1.0 indicates the program is ahead of schedule. A program with an SPI of 1.0 is exactly on schedule. The SPI has value as an "early warning indicator" when performance of a contract is declining (Abba, 2008: 29).

There are concerns, however, with SPI and its partner measurement SV (Fleming and Koppelman, 2000; Lipke, 2003). First, SV gives results in terms of dollars, and SPI is a ratio of two dollar amounts. Schedule delays are stated in dollar amounts, which seem counterintuitive. A SV of one thousand dollars could give an unclear description of how far behind schedule a program really is. Schedule performance descriptions with units of time would make more sense. Second, the mathematical calculations of SPI and SV necessitate an end result where SPI reverts to 1 and SV to zero (Fleming and Koppelman, 2000: 115, 121). Even when a program finishes late, the SPI and SV show no schedule deficiencies (Lipke, 2003: 1). The explanation is simple. The calculations for SV (BCWP-BCWS) and SPI (BCWP / BCWS) are both based on budgeted numbers instead of actual numbers. The end total of BCWS is the budget at complete (BAC). As the program approaches completion, BCWP approaches the BAC because all the work that must be completed (or is budgeted) is completed by the end of the program. Therefore, at completion, both BCWS and BCWP equal the BAC (see Figure 1 for a depiction of this). This convergence is why SV always equals zero and SPI equals 1 at the end of a program (Lipke, 2003: 3).

13
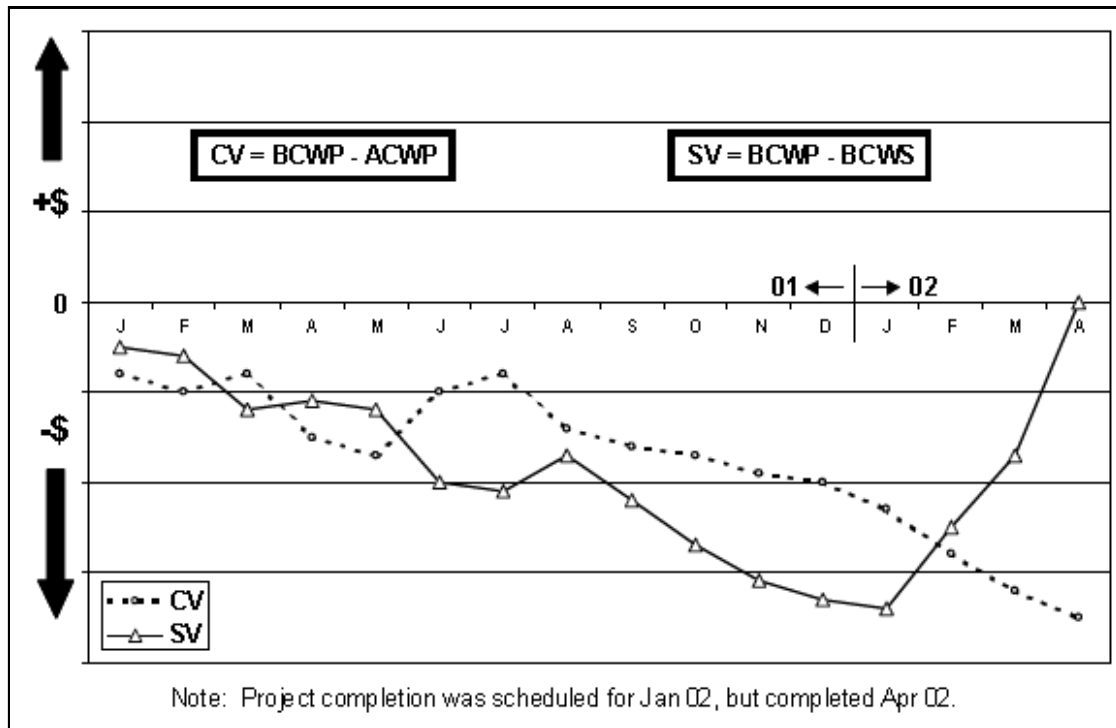
Figure 1. Cost and Schedule Variances (Lipke, 2003: 3)

Figure 1 displays an example project and displays the SV converging back to zero even though the CV decreases at the end of the project. Using this example, the SV displays a perfect performance (SV equals zero), but the graphic and note in the figure indicate the project finished three months late. Since the SV (BCWP - BCWS) equals zero at the end of the project, it means BCWP equals BCWS. Therefore, SPI equals one (BCWP / BCWS), again indicating a perfect schedule performance even though the project finished three months late. Since this SV and SPI convergence occurs, the SPI generates confusion and is not as useful once the project reaches about 67 percent complete (Lipke, 2003: 1).

SPI(t).

As a response to the reversion of SPI and SV, the concept of Earned Schedule (ES) was developed. The primary goal of earned schedule is to provide better

14

measurement of schedule performance as the program nears completion. ES's SPI(t) is

similar to SPI but is in units of time instead of money. As acknowledged by Lipke, ES

builds on from the preceding "graphical technique" of converting EVM data to a time

based variance and involves "projecting" BCWP on BCWS (Lipke, 2003: 5). Using

Figure 2 to explain the calculations involved, the first step is to identify the time

increment of BCWS with the same value as the current cumulative BCWP (in Figure 2,

the middle of June). You do this by finding the cumulative BCWP value and then going

horizontally across to the corresponding BCWS value. From there, go down vertically to

find the actual time associated with the BCWS value.
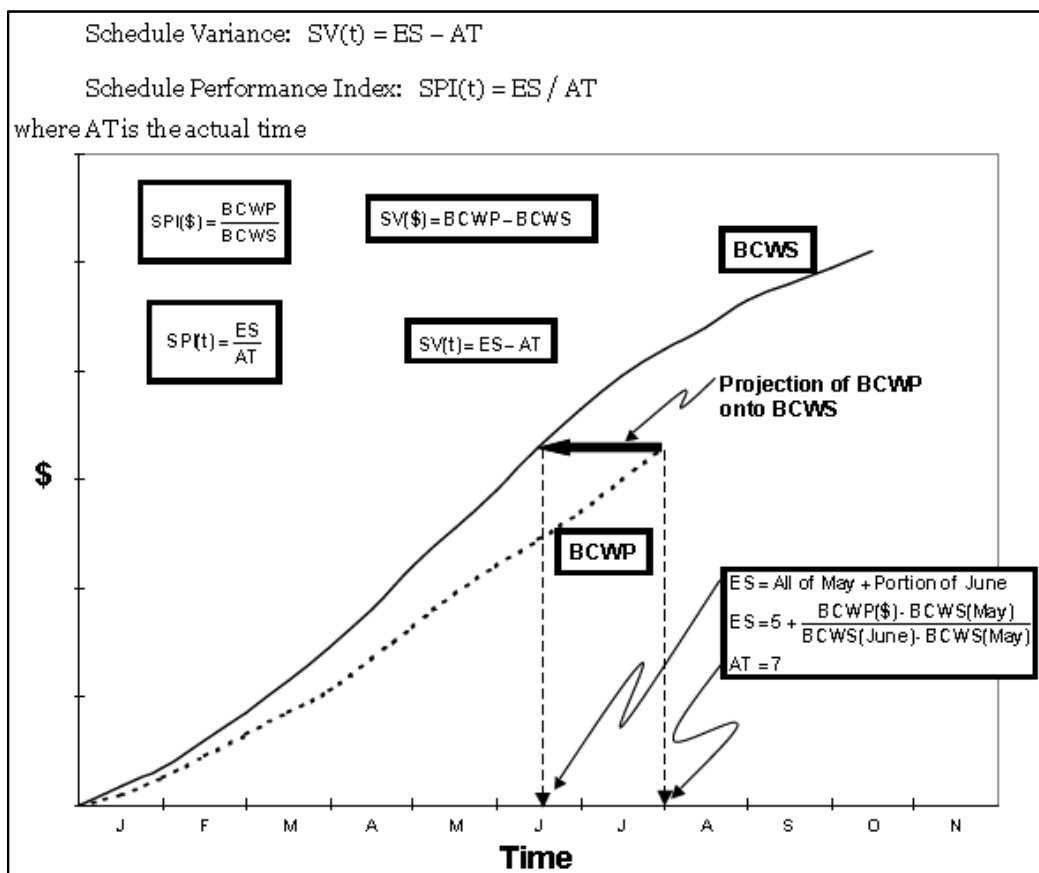
Figure 2. Earned Schedule (Lipke, 2003: 5)

ES equals the time from the start of the contract to that time increment of BCWS

(January through middle of June = 5 months + portion of June). See Figure 2 for the

formula to calculate the June portion. For simplicity in this example, the portion of June

will equal 0.5. ES equals 5.5 and the actual time (AT) is 7 months (January through

July). Therefore, the SV(t) is negative 1.5 months (5.5 - 7), and the SPI(t) is 0.79 (5.5 /

7). These calculations involve actual numbers instead of only budgeted figures as with

EVM's SV and SPI, so ES's metrics of SV(t) and SPI(t) will not converge back to 0 and

1 respectively. They remain accurate to the completion of the contract while also

providing schedule performance measurements in units of time.

The Project Management Institute's (PMI) *Earned Value Practice Standard 1st*

edition authored by a team from the then PMI College of Performance Management

(PMI-CPM) included ES as an "EVM Emerging Practice" in 2005 (PMI-CPM, 2005: 18).

ES is reported to be continually gaining interest, but actual implementation of ES in DoD

program offices currently overseeing contracts is uncommon due to the "unfamiliarity" of

the emerging practice (Crumrine, 2013: 54). Because of its accuracy over the entire

lifetime of a contract as opposed to SPI, the SPI(t) is the metric used in this research.

Past Research

This section details the past research on CPI stability and SPI(t) stability.

CPI Stability and the *Stability Rule*.

CPI stability has been discussed since the late 1970s. Thomas Bowman, a former

AFIT professor and head of the Aeronautical Systems Center's EVM department, first

heard the CPI stability rule at a semiannual C/SCSC Conference in 1978. The Air Force

Cost Center revealed the stability rule as a forecasting mechanism, and the rule stated,

"after a program is 50% complete, its cumulative CPI will not change more than plus or minus 0.10" (Bowman, 2013). In 1990, Kirk Payne performed a literature review on this CPI stability rule and concluded, "A thorough literature search revealed no published study supporting the assertion that the CPI is stable beyond the 50 percent point, although it may have been based on work done by the General Accounting Office (GAO) in the 1970s" (Christensen and Payne, 1992: 1). It was concluded that if research produced this rule, that research was never published.

The documented research into CPI stability was initially conducted by Kirk Payne in 1990. Payne, with David Christensen as his advisor, examined 26 Contract Performance Reports (CPRs) for seven different aircraft procurement contracts from the database of the Aeronautical Systems Division (ASD) (Payne, 1990: 13). Their hypothesis tested stability from the fifty percent completion point, and they used two methods to test for stability. The first method was a range of minimum and maximum CPIs from the fifty percent completion point to the end of the contract. The second method was an interval of "plus and minus 10 percent of the CPI computed at the fifty percent point" (Christensen and Payne, 1992: 2). Although the focus was to determine stability from the 50 percent completion onward, their results with the range method indicated that there was actually stability from the 20 percent completion point (Payne, 1900: 42). The interval method resulted in all the contracts being stable at the 50 percent completion point but not earlier. The result of the range method inspired the next research effort in this area of study, which sparked the current "stability rule."

The modern stability rule originated from *A Review of Cost Performance Index Stability*, by Scott Heise with guidance from Christensen (Heise, 1991). In their article,

Cost Performance Index Stability, that summarizes *A Review of Cost Performance Index Stability*, Heise and Christensen state, "Based on an analysis of 155 contracts from the DAES database, the cumulative CPI was stable from the 20 percent completion point with a 95 percent confidence interval" (Christensen and Heise, 1993: 5). The stability was defined as having a range of less than or equal to 0.20, meaning from the 20 percent completion point to the end of the program, the difference between the maximum and minimum CPI was less than or equal to 0.20 (same as the range method from Payne's research). This result became known as the "CPI stability rule." It is important to note that Christensen and Heise did not claim their finding to be a "stability rule." They did not claim generalizability to all programs that use EVM. Rather, their finding morphed into the existing stability rule of thumb that came from Payne's work and the C/SCSC conference and possible unpublished GAO study before that.

There are a couple key points to Payne's and Heise's research that are relevant to this research. First, contracts that had over target baselines (OTBs) were removed from the data set. An OTB occurs when the current budget is no longer a "realistic plan" for completing the contract, so the budget is increased (DCMA, 2006: 58). Secondly, both research efforts defined percent complete as BCWPcum / BAC (Christensen and Heise, 1993: 3; Christensen and Payne, 1992: 3). This research will use these same methods and definitions with regards to excluding contracts with OTBs and defining percent complete.

In 2002, Christensen and Carl Templin examined CPIs of 240 contracts from the DAES database. They tested a general rule of thumb, "The cumulative [CPI] will not change by more than 0.10 from its value at the 20 percent completion point, and in most cases it only worsens" (Christen and Templin, 2002: 1). To test this, they studied if the

difference between the final CPI and the CPI at 20% complete was greater than or equal to 0.10 (Christensen and Templin, 2002: 4). They stated that the results from Heise's research (the CPI stability rule) is "usually interpreted to mean" the cumulative CPI changes no more than 0.10 from the 20% completion point when in fact Heise looked to see if the range of the maximum and minimum CPI was less than or equal to 0.20. In this 2002 effort, Christensen and Templin find that the cumulative CPIs did not change by more than 0.10, "with only a few exceptions" (Christensen and Templin, 2002: 8). This study supports CPI stability, with this new, slightly different definition of stability. However, they also state that if they used the same definition as Christensen and Heise's research, it would have come up with the same result. For this research, we will look at stability with both definitions.

In the EVM community, the CPI stability rule is quoted often (Fleming and Koppelman, 2008: 17; Lipke, 2005: 14). Over time, however, the definition of the rule changed. With multiple definitions of stability (as mentioned in paragraphs above), most of the references to the stability rule interpret it differently from how it was originally stated. Table 2 lists multiple articles, starting with the seminal work on the "stability rule," that cite the stability rule along with their interpretation of the rule.

Table 2. Stability Rule Citations

Author	Stability rule interpretation
Christensen & Payne, 1992	"range of less than 0.20" and "plus or minus 10 percent of the CPI" (2)
Christensen &Heise, 1993	"range being less than 0.2" (5)
Christensen, 1996	"once a program is twenty percent completed, the cumulative CPI does not change by more than ten percent" (4)
Christensen, 1999	"cum CPI does not change by more than 10 percent from its value at 20 percent complete" (9)
Christensen & Templin, 2002 (redefined stability)	"cumulative CPI will not change by more than 0.1 from its value at the 20 percent completion point" (5)
Lipke, 2005	"the final value of the CPI does not vary by more than 0.1 from the CPI when the project is 20 percent complete" (14)
Henderson & Zwikael, 2008	"within 0.10 of its value when the project is 20 percent complete" (9)
Fleming &Koppelman, 2008	"plus or minus 10 percent" (17)
GAO, 2009	"Once a program is 20 percent complete, the cumulative CPI does not vary much from its value (less than 10 percent)" (226)
SCEA, 2010	"Cum CPI will not change more than 10% from the value at the 20% complete point in time" (124)

As demonstrated by J. Greg Smith and Kym Henderson (2008: 15), the inconsistent expression of the stability rule between plus or minus 0.10 and plus or minus 10% results in a different definition of stability as shown graphically in Figure 3. The areas between the dotted lines in the left picture and shaded in the right picture demonstrate the difference between the absolute and relative values of plus or minus 0.10.
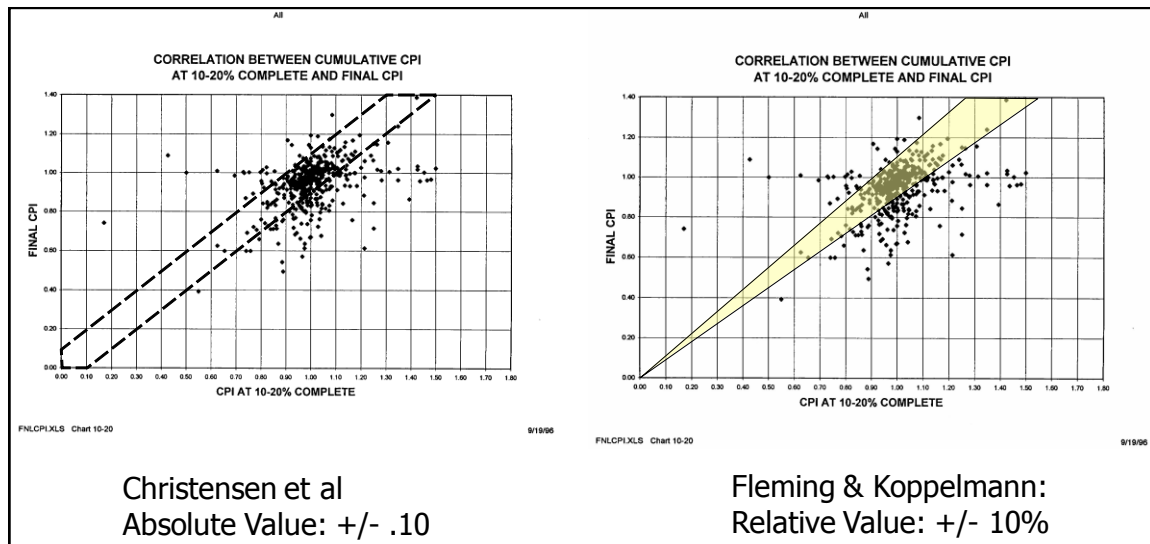
Figure 3. Comparison of CPI Stability Rule Expressed as +/- .10 and +/- 10% complete
(Smith and Henderson 2008, 15)

Along with being interpreted differently, the stability rule became a rule of thumb

that many apply to all programs using EVM. EVM handbooks and training guides quote

this stability as a rule of thumb, including the GAO *Cost Estimating and Assessment*

Guide and the Society of Cost Estimating and Analysis' (now part of the International

Cost Estimating and Analysis Association) *Cost Estimating Body of Knowledge*

(CEBoK). Some of the training guides state that rules of thumbs have exceptions

however, so they may not always be true (SCEA, 2010: 124). Kym Henderson and Ofer

Zwikael (2008) acknowledge this generalization and state that there is no evidence

supporting it. "An extensive literature review has not found further empiric validation of

the CPI stability rule beyond the project data obtained in the initial paper and data from

the … [DAES] database" (Henderson and Zwikael, 2008: 7).

Apart from the multiple interpretations shown in Table 2 and unsupported

generalization of the stability rule, another research study fifteen years after Christensen

and Heise's efforts concluded that the "widely reported CPI stability rule cannot be

generalized to all projects utilizing the EVM method or even within the DoD project portfolio," (Henderson and Zwikael, 2008: 7). Kym Henderson and Ofer Zwikael (2008) examined CPI stability using non-DoD data. Their dataset consisted of forty-five projects dealing with information technology and construction in the United Kingdom, Israel, and Australia (2008: 9). They used the Christensen and Templin definition of stability as the difference between the final cumulative CPI and the cumulative CPI at the 20% complete point being plus or minus 0.1. Their analysis concluded that the programs did not stabilize by the 20% complete point, but in fact, "the stability is usually achieved very late in the project life cycle, often later than 80 percent complete for projects in these samples" (2008: 9).

Henderson and Zwikael then turn their attention to CPI stability in DoD programs. Their analysis relies upon secondary data from Michael Popp's (1996) research on DoD projects in the U.S. Naval Air Command (NAVAIR). According to Popp's report as cited by Henderson and Zwikael, Popp's data is from the Office of the Secretary of Defense Cost Analysis Improvement Group's (OSD CAIG) Contracts Analysis System (CAS) database, a data source incorrectly thought to be different from the DAES database where Christensen and Heise gathered their data. The samples of data used in this study came from quarterly reports of the 350 development and production programs kept in the CAS (Popp, 1996: 2). The purpose of Popp's research was not to understand CPI stability but rather to "develop probability distributions of cost EACs based on the CPI at complete, current CPI, and percentage complete of projects based on history" (Henderson and Zwikael, 2008: 10). Within Popp's study, however, were charts that compared cumulative CPI to percentage complete at 10 percentile intervals. Figure 4

is one of the charts from Popp's report. Henderson and Zwikael used these charts of the

CPI data to reinforce their findings and further claim that stability could not be

generalized at the 20% complete point even for DoD programs. The dashed lines

represent the area where stability is achieved according to the stability rule utilized.

Henderson and Zwikael state that the plots outside the dashed lines are enough to

determine that the CPI stability rule does not apply for this data. Therefore, it is

"sufficient to show that the CPI stability rule cannot be generalized even within the DoD

project portfolio" (Henderson and Zwikael, 2008: 10).
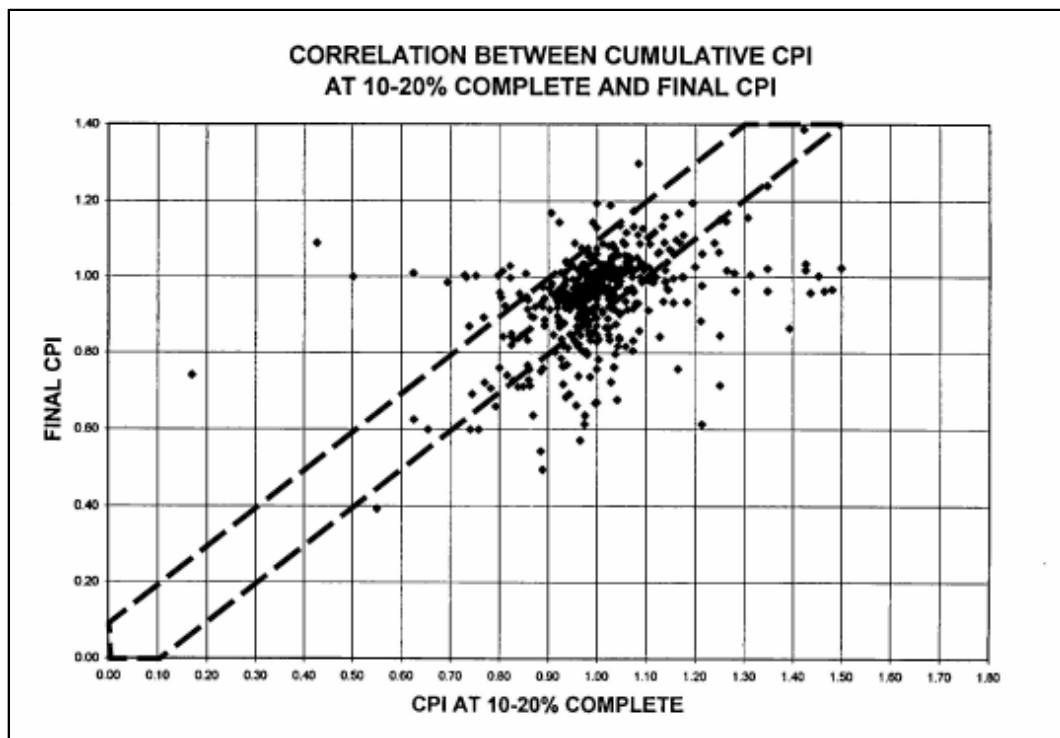
Figure 4. Correlation between Cumulative CPI at 10-20% Complete and Final CPI (Popp, 1996)

Wayne Abba challenges Henderson and Zwikael's findings in his article, *The Trouble with Earned Schedule* (2008). Abba states that Henderson and Zwikael's conclusions are wrong for several reasons. First, he notes that the secondary data from

Popp's study that Henderson and Zwikael used came from the same DAES database that Christensen and Heise also used because the CAS and DAES databases are one in the same. Abba states that Henderson and Zwikael "did not use comparable criteria to select contracts from the same source data" (Abba, 2008: 30). Second, Abba states that Henderson and Zwikael's use of primary data from the UK, Israel, and Australia cannot be compared to previous CPI stability research because there is "no evidence that these disparate projects implemented EVM consistently, as on DoD contracts," (Abba, 2008: 30). Lastly, Abba states that Henderson and Zwikael's "analysis lacks rigor. For example, Israeli data were analyzed 'using visual inspection of charts'" (2008: 30). Because of these three points, Abba argued that Henderson and Zwikael's article "does not make a case for refuting [the CPI stability rule]," (Abba, 2008: 30). While Henderson and other's subsequently contested Abba's claims, the CPI stability rule particularly within DoD has been challenged with uncertainty now prevailing.

Summarizing the literature, the CPI stability rule that Christensen and Heise confirmed has been interpreted differently and generalized by many, and it also may have been refuted by Henderson and Zwikael. This refutation was then challenged by Abba and counter challenged again. This research will use contemporary DoD data to independently see if and when the CPI displays stability characteristics to establish the ongoing validity or otherwise of the rule.

SPI(t) Stability Research in the DoD.

Although there is much research and literature about CPI stability, there is little research on possible SPI(t) stability. Henderson and Zwikael's (2008) research effort is the lone effort to date that examined SPI(t) data for stability. In contrast to the majority

of CPI stability research previously discussed, because Henderson and Zwikael could not access DoD data, they did not examine SPI(t) stability properties in DoD programs. Rather, by analyzing the forty-five information technology and construction projects in the United Kingdom, Israel, and Australia, they found the same results as with their CPI stability findings - that it was not achieved by the 20% complete point. Therefore, no conclusion was made of SPI(t) performance using DoD datasets or on large scale programs. Articles acknowledge this lack of research within the DoD (Abba, 2008: 30; Henderson and Zwikael, 2008: 8). This thesis will be the first to research the stability of SPI(t) in US DoD acquisition programs.

Importance of CPI and SPI(t) Stability

If an efficiency index stabilizes at a certain point in a contract's life, it would bring many benefits to the PM. First, it provides a better forecasting or estimating ability (Bowman, 2013). For example, CPI is part of the calculation for the EAC. Therefore a stable CPI would provide better EACs because of the reduced variation. The better forecasting also allows PMs to avoid the mistake of committing more funds to a failing project, which can be very costly to all stakeholders involved with that program (Christensen, 1996). Secondly, the CPI serves as a "benchmark" to the To Complete Performance Index (TCPI), which displays the CPI needed for the rest of the contract's life in order to meet the BAC (Christensen and Heise, 1993: 2). If the TCPI is much higher than the CPI, the TCPI may be unattainable since the CPI is stable. Therefore, the program will not be able to reach the BAC and will most likely finish over budget. CPI stability gives confidence to this conclusion. Thirdly, a stable performance index is claimed to be evidence that the contractor's management system is working. The

performance is stable, meaning there are no large variances unaccounted for or unaddressed by the management of that contract (Christensen and Heise, 1993: 2). While the CPI is the subject of all these referenced benefits, stability in either of the efficiency indices may provide these benefits since the PM uses both cost and schedule to manage a program.

Many training guides and handbooks used in the EVM community quote the CPI stability rule. The quotes are included in Table 2 and mentioned earlier in this chapter. Key documents that mention CPI stability include the CEBoK and the GAO Cost Estimating and Assessment Guide. This fact attests to the argument that this stability is important and very useful.

Conclusion

EVM is a critical tool in a PM's toolbox and leans heavily on the use of the CPI and SPI. As stated by Fleming and Koppelman, the CPI is perhaps the most important measure in EVM (2008). The CPIs "stability rule" may be used by many in the EVM community, but it has been interpreted in different ways and incorrectly generalized (as seen in Table 2). The stability definition changes from between a range of 0.2 to plus or minus 0.10 and plus or minus 10 percent, as well as said to exist among all programs using EVM within or external to DoD. This thesis will examine cumulative CPIs of DoD acquisition programs to determine if stability exists, using the definitions of stability outlined in the next chapter. The other efficiency index, SPI, is flawed because of its inherent convergence to one and sometimes confusing because of its measure in monetary units. ES's SPI(t) does not possess these same faults and is becoming a more

prominent measure to monitor schedule performance of a program. There is no in-depth

research into SPI(t) stability within DoD programs. This research project will be the first.

Now that all necessary terms involving PM, EVM, CPI, and SPI(t) have been

defined and all research into the CPI stability rule and SPI(t) possible stability has been

addressed, the next step is to walkthrough the methodology of this research.

III. Methodology

Overview

Chapter 3 provides detailed information about the data and the methodology utilized to analyze the data. First, it explains the source of the data and the data collection process, including decisions on what data to include or remove. Then, we define important formulas used in the dataset to prepare it for analysis. Next, this chapter explains the variance analysis of this research with hypothetical examples to facilitate understanding of the three different analyses of measuring stability. Finally, we outline the hypothesis testing we utilize in our comparison analyses between services, contract types, lifecycle phases, and platforms.

Data

The dataset used in the research is from the Defense Acquisition Executive Summary (DAES) and Defense Cost and Resources Center (DCARC) databases. Both sources are used to include as much data as possible. DCARC'S EVM Central Repository (which was directed in July 2007 by USD (AT&L) to be utilized as the central repository for all ACAT I EVM data reporting) contains modern data (up to current date) (USD-AT&L, 2008: 10). Data for the earlier years of the analysis, 1987 through 2002, are from DAES. The complete dataset consists of reported EVM numbers at the contract level (the ACWP, BCWP, BCWS, and BAC) from all DoD Acquisition Category 1 (ACAT 1) programs from 1987 to 2012 with available EVM data. We include any ACAT 1 program's contract that is no more than 10% complete by the start of 1987 and at least 85% complete by the end of 2012 (percent complete is defined in the Percent Complete section of this chapter). This required range of percent complete

allows only contracts with complete data into the dataset, which is necessary for our

analysis where we examine the entire life of the contract. We chose 85% complete to be

equivalent with the final report of the contract to be more conservative than past research

that utilized an 80% completion point (Christensen and Templin, 2002). There is some

evidence, however, that the true equivalency completion point is closer to the final

report's 92.5% complete point (Tracy and White, 2011). But to include as many

contracts in this study as possible, we assume 85% complete to be equivalent to the final

report of the contract. This results in thirty-four contracts that do not reach 92.5%

complete but are retained in the dataset because they meet our 85% complete

requirement.

Consistent with previous research (Heise, 1991; Christensen and Templin, 2002),

we also remove from the dataset all contracts with Over-Target-Baselines (OTBs). With

a re-baseline, the budget of a contract changes. This change causes all the EVM

budgeted measurements to restart, including the cumulative measurements. See Table 3

for a summary of the combined DAES and DCARC dataset before and after the OTBs

were removed and the time window/percent complete requirements were enforced.

Table 3. Dataset Characteristics

Dataset Characteristics	
Category	**Total Contracts**
Preliminary Total (before percent complete and OTB cuts)	822
Contracts with OTBs	165
Does not meet the time window/percent complete requirements	447
Final Dataset for analysis	209
Service	
Army	45
Navy	97
Air Force	65
Life-cycle Phase	
Production	102
Development	102

The final dataset includes 209 contracts. Contracts include AF, Navy, and Army contracts from all life-cycle phases. Table 3 provides a summary of the data. See Appendix B for a complete listing of the programs included in the analysis.

Adding Formulas to the Data

After collecting all the relevant contracts, we organize the data in preparation for analysis. For each data point, there are four main EVM measurements we use: Budgeted cost for work performed (BCWP), Budgeted cost for work scheduled (BCWS), actual cost of work performed (ACWP), Budget at Complete (BAC). Explanations of these measurements can be found in Chapter 2. BCWP, ACWP, and BCWS are all cumulative. All subsequent references to these measurements refer to the cumulative version of the measurement.

Percent Complete.

The formula for Percent Complete used in this research is BCWPcum / BAC. The BAC in this formula is the final BAC of the contract. The BAC often changes throughout the contract's life. If the BAC increases a substantial amount from one month to the next, the percent complete could actually decrease if the BAC increase is larger than the increase in BCWPcum. This generates inconsistent results. Therefore, in order to have a stable denominator and ensure the percent complete moves in one direction throughout the entire life of the contract, we use the final BAC (the BAC of the last data point for each contract) as the denominator for all percent complete calculations of that contract. This approach is consistent with previous research efforts including Christensen and Payne (1992).

For our analysis, we analyze the CPI and SPI(t) of each contract in 5 percent increments from 10% complete to 85% complete. Most percent complete calculations, however, do not result in exact 5 percent increments, such as 20% or 35%. Instead, they result in decimals. To find a specific percent complete point in a contract's life, we define any percent complete within 2.5% of the specific percent complete to be that specific percent complete point. For example, if examining the CPI at 20% complete and we have CPIs from when the contract is 18% complete (0.86) and 24% complete (0.90), we use the 18% complete measurements for the CPI at 20% complete. The CPI at 24% complete is more than 2.5% from 20% so it is not used in calculating the CPI at 20%. If more than one point is within 2.5% of the specific increment of five percent complete, we average them to determine the CPI or SPI(t) for that five percent complete increment.

We use this process for finding all CPIs and SPI(t)s at exact percent completes during the variance analysis part of this research.

CPI.

We calculate the CPI using the formula given in Chapter 2: BCWPcum / ACWPcum. Both of these numbers are the cumulative measurements since we are analyzing the cumulative CPI. Throughout this chapter, "CPI" refers to the cumulative CPI.

SPI(t).

The SPI(t) is calculated using the Earned Schedule (ES) calculator titled "ES Calculator v1b" found on the ES website (http://www.earnedschedule.com/Calculator.shtml). To use the ES calculator, BCWP and BCWS values are put into the spreadsheet, and the resulting cumulative SPI(t) for each data point is displayed. Throughout this chapter, "SPI(t)" refers to the cumulative SPI(t).

Variance Analysis

With all the CPIs, SPI(t)s, and percent completes calculated, the next step is to execute variance analysis on the CPI and SPI(t) to see when each stabilizes in every contract. Variance analysis will be executed on a contract's CPI and SPI(t) from the 10 percent complete point to 85 percent complete in increments of five. We complete three separate analyses, one analysis for each definition of stability.

First Analysis: Range Definition of Stability.

First, we define stability as when the difference between the maximum and minimum CPI or SPI(t) between a specific percent completion and the final point is less

than 0.2. We chose 0.2 as the range to follow Christensen, Payne, and Heise's works as discussed in Chapter 2 (Christensen and Payne, 1991; Christensen and Heise, 1993). The test for this definition is:

$$\text{Stable if: } |CPImax_{X\%} - CPImin_{X\%}| \leq 0.20$$

$$\text{Unstable if: } |CPImax_{X\%} - CPImin_{X\%}| > 0.20$$

$CPImax_{X\%}$ equals the maximum CPI from the X% complete point to the final CPI of the contract, and $CPImin_{X\%}$ equals the minimum CPI from the X% complete point to the final CPI of the contract. We define X as a percent complete from 10 to 85 percent complete in increments of five. The formula is for the CPI, but we use the same formula for SPI(t) stability by replacing CPI with SPI(t). We find if the contract is stable and record the calculated range at each percent complete increment.

To better understand this first definition and analysis, Figure 5 has CPI measurements from the simulated contract with the Contract Identity Description (CID) "ABCD123." For simplicity, this example has CPI data from 20% complete to the completion of the contract (in this case, 95% complete is the contract's final report). The graph below the data in Figure 5 shows the trend of the CPI along the life of the contract. Looking for stability from the 20% complete point, we find the maximum and minimum CPIs between the 20% point and the 95% point and calculate the difference between the two CPIs. In this example, the maximum is 1.03 (at the 40% complete point), and the minimum is 0.87 (at the 90% complete point). The difference between these two CPIs is 0.16, which is less than 0.2, so this contract has stability as we defined it earlier in this paragraph. For simplicity in this example, the 10% and 15% complete points were not

included. However, since this contract is stable at the 20% complete point, we would

look at these earlier points to see if the CPI is stable starting earlier than at 20% complete.

	Percent Complete Point															
CID-ABCD123	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95
CPI	0.9	0.94	0.96	1.01	1.03	1.01	1.01	0.98	0.97	0.95	0.92	0.9	0.88	0.89	0.87	0.89
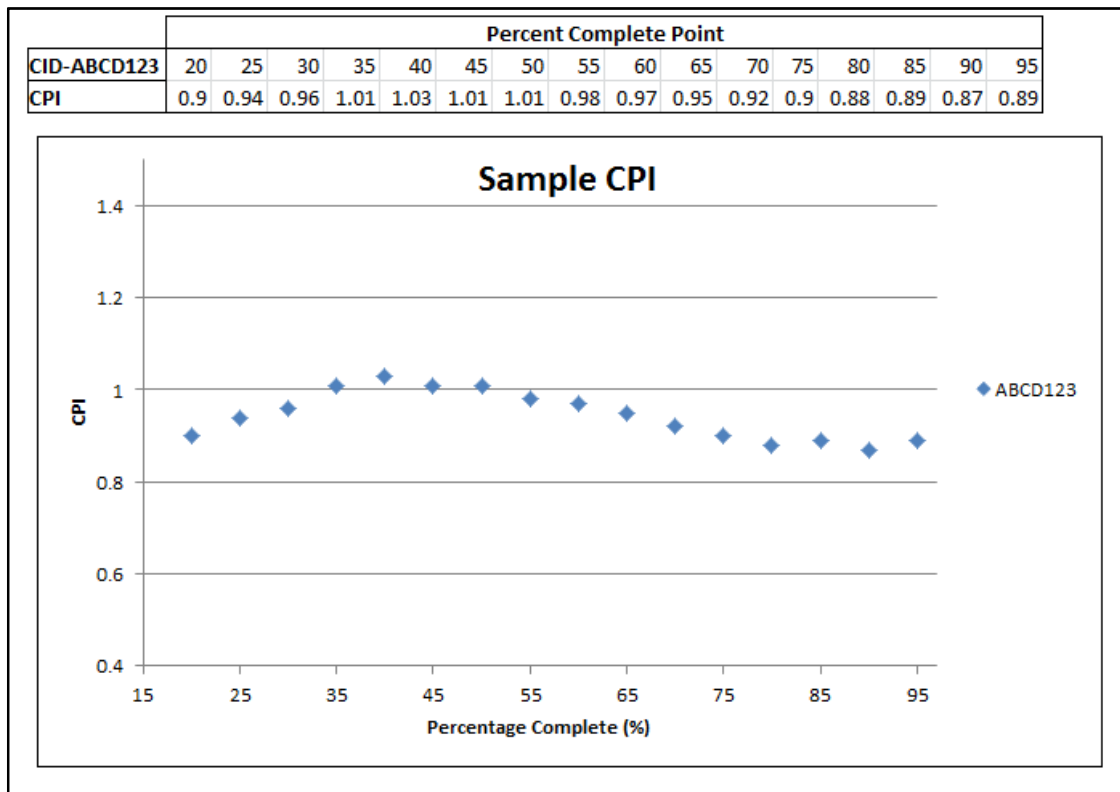
Figure 5. Sample of CPI stability

If the difference between the minimum and maximum CPIs from the 20%

completion point to the end point is larger than 0.2, we examine the difference between

the minimum and maximum CPIs from the 25% completion point to the end point and

then from the 30% completion point to the end point. We continue this process of

finding differences until we find a point where stability is achieved, where the difference

is less than 0.2. We also record the range for each increment of percent complete.

Once that range between the minimum and maximum CPIs (or SPI(t)s) is found

for each contract, we construct a confidence interval to determine the mean range of all

contracts for each percent complete increment. The confidence interval will better describe the statistics of CPI and SPI(t) stability.

Second Analysis: Absolute Interval Definition of Stability.

For the second analysis, we define stability as when the final CPI (or SPI(t)) is within 0.10 of the CPI (or SPI(t)) at a specific percent complete. The test for this analysis is:

$$\text{Stable if: } |CPI(final) - CPI(X\%)| \leq 0.10$$

$$\text{Unstable if: } |CPI(final) - CPI(X\%)| > 0.10$$

$CPI(final)$ is the final CPI of the contract, and $CPI(X\%)$ is the CPI when the contract is $X\%$ complete. We define X as a percent complete from 10 to 85 in increments of five. For example, the contract is stable from 20% complete if the difference between the final CPI and the CPI at 20% complete is less than plus or minus 0.10 (for this example, X equals 20%). If the difference is greater than 0.10, it is not stable at 20% complete. When testing SPI(t), we use the same formula by replacing the CPI with SPI(t). We examine all percent complete increments from 10% to 85% complete to see if each efficiency index is stable at each. As with the first analysis, we record the differences at every percent complete increment of five for all the contracts and create a confidence interval around each of these.

Third Analysis: Relative Interval Definition of Stability.

For the third analysis, we define stability as when the difference between the final CPI and the CPI of a specific percent complete is less than or equal to plus or minus 10% of the CPI at the specific percent complete. For instance, a contract with a CPI at 20%

complete of 0.9 stabilizes from that point if the final CPI is less than or equal to plus or minus 0.09 (ten percent of 0.9). The test is:

$$\text{Stable if: } |CPI(final) - CPI(X\%)| \leq 0.10 \cdot CPI(X\%)$$

$$\text{Unstable if: } |CPI(final) - CPI(X\%)| > 0.10 \cdot CPI(X\%)$$

$CPI(final)$ is the final CPI of the contract, and $CPI(X\%)$ is the CPI when the contract is $X\%$ complete. We define X as a percent complete from 10 to 85 in increments of five. When testing SPI(t), we use the same formula by replacing the CPI with SPI(t).

As with the other two analyses, we record the differences at every percent complete increment of five from all the contracts and create a confidence interval around each of these.

Comparison Analysis

Once we find the stability points for each analysis, we categorize contracts and compare the differences in the CPI ranges (from the first stability analysis) and intervals (from the second stability analysis), as well as the ranges and intervals of SPI(t). We execute four different comparisons: service, contract type, acquisition life-cycle phase, and platform. For service, we compare Army, Navy, and Air Force contracts to see if one service's contracts' CPI and SPI(t) tend to stabilize earlier than others. See Table 4 for a list of the different services, contract types, life-cycle phases, and platforms we compare.

Table 4. Categories for Comparison Analysis

Categories			
Services	**Contract Types**	**Life-cycle Phases**	**Platforms**
Air Force	Fixed Price	Development	Aircraft System
Army	Cost Plus	Production	Electronic/Automated System
Navy			Missile System
			Ordnance System
			Ship System
			Space System
			Surface Vehicle System

For each of these categorical comparisons, we utilize the same hypothesis test:

$$\text{Ho: } \Delta_X = \Delta_Y$$

$$\text{Ha: } \Delta_X \neq \Delta_Y$$

Δ equals the median of the ranges (from the first stability analysis) or interval (from the second stability analysis) for the specific percent complete (values of 10% to 85% , in increments of five). X and Y are different groups in each comparison. For example, when comparing services, we define X and Y as Air Force and Navy, Air Force and Army, and Navy and Army respectively (three different tests). If we fail to reject the null, there is no difference in the median range or interval of the groups X and Y. If we reject the null hypothesis, it means there is a difference in the medians. We state which is larger according to the results.

Statistical Tests.

For our hypothesis testing, we utilize the Kruskal-Wallis and Mann-Whitney tests to compare the medians. We cannot confidently perform t-tests to compare the mean values, as the variables have many outliers which make the assumption of normality difficult to justify. For the Service comparison and Platform comparison, we require the

Kruskal-Wallis test. The Kruskal-Wallis test is the overall test we utilize when there are more than two groups under one category. If it is significant, we then execute the Mann-Whitney test to compare each pair of groups. For Life-cycle Phase and Contract Type, there are only two groups. Therefore, we just perform Mann-Whitney tests. We execute each test using the range values from the range definition and interval values from the absolute interval definition. An explanation of the tests follows.

Kruskal-Wallis.

The Kruskal-Wallis test is nonparametric, meaning it makes no assumption on distribution (Hayter, 2007: 713). The null hypothesis is that the medians of the three or more groups, or variables, it is comparing are equal (for example, the service comparison has three variables to compare: Air Force, Navy, and Army ranges or intervals). Therefore, it is only used when comparing the Service and Platform categories. It is a rank sum test, so it combines all the observations for all the variables and ranks them. The sum of the ranks is

$$1 + \cdots + n_T = \frac{n_T(n_T + 1)}{2}$$

where n_T is the total number of observations (Hayter, 2007:713). Next, the test calculates the rank averages within each variable. This is the formula for the rank average:

$$\bar{r}_{i.} = \frac{r_{i1} + \cdots + r_{in}}{n}$$

$r_{i.}$ is the i-th rank, and n is the total number of observations for the variable that the i-th rank belongs to. Using these rank averages, the next step calculates the test statistic H, using this formula (Hayter, 2007:713):

$$H = \frac{12}{n_T(n_T + 1)} \sum_{i=1}^{k} n_i \left(\bar{r}_{i.} - \frac{(n_T + 1)}{2} \right)^2$$

The p-value of this test (gained using H) will be compared to the family-wise error rate (α_e). α_e is 0.05, the level of significance of the hypothesis. If the test results are significant (we reject that the medians are equal), we then execute Mann-Whitney tests to compare the possible pairs among that category. These secondary tests will be at the level of significance equal to the comparison-wise error rate (α_c), which is α_e divided by the number of secondary tests in accordance with statistical theory on multiple simultaneous comparisons and error rates known as the Bonferroni method (Kutner, 2004).

Mann-Whitney.

The Mann-Whitney test is also known as the Wilcoxon test. It is a "distribution-free" test, so there is no assumption on the normality of the distribution (Hogg and Craig, 1995: 498). This test compares the median of the two variables, or groups. For the test, we calculate the test statistic, U, using this formula (Hogg and Craig, 1995: 522):

$$U = \sum_{j=1}^{n} \sum_{i=1}^{m} Z_{ij}$$

m is the number of Ys, and n is the number of Xs. X and Y are the two variables or groups we are comparing. Z_{ij} equals one if X_i is less than Y_j, and equals zero if X_i is greater than Y_j. The summation formula counts the total number of X values that are less than each value of Y (Hogg and Craig, 1995: 522). We compare the resulting U to the critical values in a significance table. We use a level of significance (α) of 0.05 for the Contract Type and Life-cycle Phase comparisons because there are just two variables in

each comparison. Since there are just two variables in these comparisons, the Mann-Whitney test serves as the overall test and there is no need for the initial Kruskal-Wallis test. For the Kruskal-Wallis tests for Service and Platform, we use alpha of 0.05. In the follow-on Mann-Whitney tests for Service, we use an alpha of 0.0167 (0.05 divided by 3), and for Platform we use an alpha of 0.0071 (0.05 divided by 7).

Conclusion

This chapter reviews the data collection process and describes the resulting dataset. We use EVM data of 209 contracts from 1987 to 2012 from the DAES and DCARC databases. We calculate the CPI and SPI(t) of each data point. Using three different definitions of stability, we find the point where stability is achieved for each measurement (CPI and SPI(t)) in each contract. Once a stability point is found for each contract, we provide descriptive statistics on the stabilities. Then, we compare the military services, contract types, life-cycle phases, and platforms to see if a certain group's ranges (from the range definition of stability) or intervals (from the absolute interval definition) are different, utilizing the Mann-Whitney and Kruskal-Wallis tests. Chapter 4 describes our results from this methodology.

IV. Results

Overview

This chapter provides the results from the methodology outlined in Chapter 3. We examine the results of CPI stability first and then SPI(t) stability. First, descriptive statistics and confidence intervals provide the results of the variance analysis from the three different definitions of stability. Then, the results of the individual hypothesis testing determine the comparisons in stability characteristics between services, contract types, acquisition life-cycle phases, and platforms.

CPI Variance Analysis

First, we study if and when the CPI stabilizes for each contract. We look at stability using the three different definitions explained in Chapter 3. For each analysis of stability, descriptive statistics show what percentage of contracts stabilizes from each percent complete point. A confidence interval provides an estimate for the mean range or interval (depending on the analysis of stability) of the contracts.

First Analysis: Range Definition of Stability.

The first analysis of stability defines stability as when the difference between the maximum and minimum CPI (or the range) is less than or equal to 0.20. Table 5 displays the results of the first analysis of CPI. The last two rows of the table are the upper and lower bounds of the 95% confidence interval, which focuses on the actual value of the calculated ranges. The t-critical value is 1.971 for all the intervals since each has the same degrees of freedom, 208.

Table 5. Range Stability of CPI

CPI Analysis																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Number of contracts	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209
Number of stable contracts	152	164	173	175	183	191	195	197	199	201	205	206	206	207	208	208
Percentage of stable contracts	72.7	78.5	82.8	83.7	87.6	91.4	93.3	94.3	95.2	96.2	98.1	98.6	98.6	99.0	99.5	99.5
Mean Range	0.17	0.15	0.13	0.12	0.11	0.10	0.09	0.08	0.08	0.07	0.06	0.05	0.05	0.04	0.03	0.02
Range Std Deviation	0.14	0.12	0.10	0.08	0.08	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.03	0.03
Upper Confidence Limit	0.19	0.16	0.14	0.13	0.12	0.11	0.10	0.09	0.08	0.08	0.07	0.06	0.05	0.04	0.03	0.03
Lower Confidence Limit	0.15	0.13	0.12	0.11	0.10	0.09	0.08	0.07	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.02

Second Analysis: Absolute Interval Definition of Stability.

For the second analysis, stability is when the difference between the CPI at a specific percent complete and the final CPI (or the interval) is less than or equal to 0.10. Therefore, the contract's CPI is stable if the final CPI is within an interval of plus or minus 0.10 of the CPI at the specific percent complete. Table 6 summarizes the results of the second analysis of CPI stability. There are different amounts of contracts for each percent complete increment because some contracts do not have a data point in a certain percent complete increment.

Table 6. Absolute Interval Stability of CPI

CPI Analysis																
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Number of contracts	173	173	149	149	147	149	137	137	153	147	145	155	149	164	171	174
Number of stable contracts	103	106	99	106	101	111	104	102	127	119	127	134	132	149	164	172
Percentage of stable contracts	59.5	61.3	66.4	71.1	68.7	74.5	75.9	74.5	83.0	81.0	87.6	86.5	88.6	90.9	95.9	98.9
Mean Interval	0.13	0.11	0.10	0.09	0.08	0.07	0.07	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.02
Interval Std Deviation	0.13	0.12	0.10	0.09	0.08	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03
t-critical value	1.97	1.97	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.97	1.97	1.97
Upper Confidence Limit	0.15	0.13	0.11	0.10	0.10	0.09	0.09	0.08	0.07	0.07	0.06	0.05	0.05	0.04	0.03	0.03
Lower Confidence Limit	0.11	0.10	0.08	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.02	0.02

When examining the contracts with this second definition of stability, however, we discover an issue. The equation only accounts for the single CPI at a specific percent complete and then the final CPI. Therefore, the contract may be stable from the 10% complete point because the CPI at 10% complete is within 0.10 of the final CPI. But this same contract can be unstable from the 20% complete point if the CPI at 20% complete is not within 0.10 of the final CPI. The CPI at 20% complete does not affect the stability from the 10% complete point. Table 7 illustrates this issue with a hypothetical contract ABCD123. The table provides the CPI for each percent complete increment, along with the difference between it and the final CPI. The last row displays whether or not the contract is stable at that increment (S = stable, U = unstable). The interval and stability cells for the last three percent completes are blank because we only calculate stabilities

up to the 85% complete increment. The CPI is stable at 10% complete, with a CPI equal to the final CPI. However, the contract is unstable at the 15% and 20% complete increments (both have differences greater than 0.10). Therefore, it is misleading to say the contract is stable from the 10% complete point when it loses that stability at times after that percent complete.

Table 7. Issue with Second Analysis

Contract ABCD123	Percent Complete																		
	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
CPI	1.07	0.93	0.93	NA	1.00	0.93	0.92	0.95	1.02	1.05	1.06	1.06	1.06	1.07	1.07	NA	1.07	1.07	1.07
Interval	0.00	0.14	0.14	NA	0.07	0.14	0.15	0.12	0.05	0.02	0.01	0.01	0.01	0.00	0.00	NA	-	-	-
Stability	S	U	U	NA	S	U	U	U	S	S	S	S	S	S	S	NA	-	-	-

This issue of changing from stable to unstable could happen more than once in a contract. When examining the CPI, 46 contracts exhibit this issue once, 18 exhibit it twice, and one exhibits it three times. With SPI(t), 44 contracts exhibit this issue once, 19 twice, and 1 three times. The issue is when, once the contract is stable, it becomes unstable before the end of the contract. Twice means it becomes stable, then unstable, back to stable, and changes to unstable once more (changes to unstable two times before end of contract). To address this issue, we adjust the stability definition, so the contract is only stable if the difference of the final CPI and specific CPI is less than or equal to 0.10 and *maintains* that 0.10 interval throughout the rest of the contract from that specific percent complete. This is consistent with past research (Christensen and Payne, 1992; Henderson and Zwikael, 2008). To address missing data points, we assume a five percent increment that has no data is stable if it is surrounded by two increments that are stable or it is the last five percent increment and the one before it is stable. In our example in Table 7, the 25% complete and 85% complete increments do not have data.

44

We consider the 25% increment unstable since it is not surrounded by stable increments (the 20% increment is not stable). The contract is stable at the 85% complete increment, however, since it is the last increment and the increment right before it (80%) is stable. Since we count an increment with no data as stable or unstable, there are 209 total contracts for each increment. This total is the denominator when calculating the percentage of stable contracts. Table 8 displays the adjusted results from addressing this issue. There is no different confidence interval for these results since the intervals stay the same.

Table 8. Absolute Interval Stability Adjusted Results of CPI

CPI Analysis															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Number of stable contracts	83	106	117	125	129	139	146	151	159	169	175	179	186	192	200	206
Percentage of stable contracts	39.7	50.7	56.0	59.8	61.7	66.5	69.9	72.2	76.1	80.9	83.7	85.6	89.0	91.9	95.7	98.6

Third Analysis: Relative Interval Definition of Stability.

The third analysis defines stability as when the final CPI is within a relative interval of the CPI at the specific percent complete. The difference between the final CPI and the CPI at a specific percent complete must be less than or equal to ten percent of the CPI at the specific percent complete. Table 9 provides the results of the third analysis of CPI stability. The third analysis results in the same confidence interval as the second analysis because the confidence interval is over the same difference, between the CPI at the certain percent complete and the final CPI. Therefore, it is not included in the table.

Table 9. Relative Interval Stability of CPI

CPI Analysis																
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Number of contracts	173	173	149	149	147	149	137	137	153	148	145	155	149	164	171	174
Number of stable contracts	102	108	102	103	105	110	104	101	124	118	123	132	132	151	163	171
Percentage of stable contracts	59.0	62.4	68.5	69.1	71.4	73.8	75.9	73.7	81.0	80.3	84.8	85.2	88.6	92.1	95.3	98.3

Similar to the second analysis, contracts can be stable from an early percent complete but be unstable from a later percent complete. With the CPI, 34 contracts exhibit this issue once, five exhibit it twice, and one exhibits it three times. With the SPI(t), 56 exhibit the issue once, and 14 exhibit it twice. Therefore, we re-examine the contracts and only consider the contracts *stable if they remain stable from that point onward*. Following the same assumptions as with the second analysis, each percent complete increment for the adjusted results contains 209 contracts (which is the denominator when calculating the percentages of stable contracts) since an increment with no data is stable if it is surrounded by increments that are stable or is the final increment and the one before it is stable. Table 10 shows the adjusted results.

Table 10. Relative Interval Stability Adjusted Results of CPI

CPI Analysis																
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Number of stable contracts	83	106	115	121	126	134	141	145	152	162	171	175	181	189	197	204
Percentage of stable contracts	39.7	50.7	55.0	57.9	60.3	64.1	67.5	69.4	72.7	77.5	81.8	83.7	86.6	90.4	94.3	97.6

Summary of CPI Stability.

Table 11 provides a summary of the three analyses executed on CPI stability- the Range stability, Absolute (Abs) Interval stability, and Relative (Rel) Interval stability. Bolded are the points where the percentage of stable contracts initially reaches 70, 80, and 90%. If a rule of thumb or heuristic is to be drawn from these results, the individual/organization using this heuristic must determine the relevant percentage: 70%, 80%, 90%, or some other percentage.

Table 11. Summary of CPI Stability

CPI Summary- Percentage of stable contracts															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Range Stability	**72.7**	78.5	**82.8**	83.7	87.6	**91.4**	93.3	94.3	95.2	96.2	98.1	98.6	98.6	99.0	99.5	99.5
Abs Interval Stability	39.7	50.7	56.0	59.8	61.7	66.5	69.9	**72.2**	76.1	**80.9**	83.7	85.6	89.0	**91.9**	95.7	98.6
Rel Interval Stability	39.7	50.7	55.0	57.9	60.3	64.1	67.5	69.4	**72.7**	77.5	**81.8**	83.7	86.6	**90.4**	94.3	97.6

SPI(t) Variance Analysis

We recreate the same methodology using SPI(t) instead of CPI. We look at SPI(t) stability using the three different definitions explained in Chapter 3. For each analysis of stability, descriptive statistics show what percentage of contracts stabilizes from each percent complete point. A confidence interval provides an estimate for the mean range or interval (depending on the analysis of stability) of the contracts.

First Analysis: Range Definition of Stability.

The first analysis of stability defines stability as when the difference between the maximum and minimum SPI(t) (or the range) is less than or equal to 0.20. Table 12

provides the results. The last two rows of the table are the upper and lower bounds of the

95% confidence interval, which focuses on the actual value of the calculated ranges. The

t-critical value is 1.971 for all the intervals since each has the same degrees of freedom,

208.

Table 12. Range Stability of SPI(t)

SPI(t) Analysis																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Number of contracts	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209
Number of stable contracts	150	163	171	173	174	180	184	185	187	189	191	193	197	199	201	203
Percentage of stable contracts	71.8	78.0	81.8	82.8	83.3	86.1	88.0	88.5	89.5	90.4	91.4	92.3	94.3	95.2	96.2	97.1
Mean Range	0.17	0.15	0.14	0.13	0.12	0.12	0.11	0.10	0.10	0.10	0.08	0.08	0.07	0.06	0.05	0.04
Range Std Deviation	0.17	0.16	0.15	0.15	0.15	0.14	0.14	0.14	0.14	0.14	0.10	0.08	0.08	0.07	0.06	0.05
Upper Confidence Limit	0.19	0.17	0.16	0.15	0.14	0.14	0.13	0.12	0.12	0.11	0.10	0.09	0.08	0.07	0.06	0.05
Lower Confidence Limit	0.14	0.13	0.12	0.11	0.10	0.10	0.09	0.09	0.08	0.08	0.07	0.06	0.06	0.05	0.04	0.03

Second Analysis: Absolute Interval Definition of Stability.

For the second analysis, stability is when the difference between the SPI(t) at a

specific percent complete and the final SPI(t) is less than or equal to 0.10. Therefore, the

contract is stable if the final SPI(t) is within an interval of plus or minus 0.10 of the SPI(t)

at a specific percent complete. Table 13 displays the results from the second analysis of

SPI(t).

Table 13. Absolute Interval Stability of SPI(t)

SPI(t) Analysis																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Number of contracts	173	173	149	149	147	149	137	137	153	147	145	155	149	164	171	174
Number of stable contracts	124	128	117	113	115	118	119	107	129	121	118	128	123	139	152	155
Percentage of stable contracts	71.7	74.0	78.5	75.8	78.2	79.2	86.9	78.1	84.3	82.3	81.4	82.6	82.6	84.8	88.9	89.1
Mean Interval	0.09	0.08	0.07	0.07	0.07	0.07	0.05	0.06	0.05	0.07	0.06	0.05	0.05	0.05	0.04	0.04
Interval Std Deviation	0.11	0.11	0.10	0.10	0.09	0.08	0.07	0.08	0.07	0.14	0.09	0.06	0.07	0.07	0.05	0.05
t-critical value	1.97	1.97	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.98	1.97	1.97	1.97
Upper Confidence Limit	0.11	0.10	0.08	0.09	0.08	0.08	0.06	0.08	0.06	0.09	0.07	0.06	0.06	0.06	0.05	0.04
Lower Confidence Limit	0.07	0.07	0.05	0.06	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.04	0.03	0.03

As mentioned previously when examining CPI, the second analysis allows contracts to be stable from an early percent complete but unstable from a later percent complete. To address this issue, we re-examine the stabilities and only consider a contract stable if it remains stable from that point onward. Once again, we assume an increment with no data is stable if it is surrounded by increments that are stable or is the last increment and the one before it is stable. This means there are a total of 209 contracts for each increment. Table 14 shows the adjusted results.

Table 14. Absolute Interval Stability Adjusted Results of SPI(t)

	SPI(t) Analysis															
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Number of stable contracts	86	109	122	125	130	139	142	144	147	152	157	164	167	175	182	186
Percentage of stable contracts	41.1	52.2	58.4	59.8	62.2	66.5	67.9	68.9	70.3	72.7	75.1	78.5	79.9	83.7	87.1	89.0

Third Analysis: Relative Interval Definition of Stability.

The third analysis defines stability as when the final SPI(t) is within a relative interval of the SPI(t) at a specific percent complete. The difference between the final SPI(t) and the SPI(t) at a specific percent complete must be less than or equal to ten percent of the SPI(t) at the specific percent complete. Table 15 summarizes the results of the third analysis. The third analysis results in the same confidence interval as the second analysis because the confidence interval is over the same difference, between the SPI(t) at the specific percent complete and the final SPI(t). Therefore, the confidence interval is not included in the table.

Table 15. Relative Interval Stability of SPI(t)

	SPI(t) Analysis															
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Number of contracts	173	173	149	149	147	149	137	137	153	147	145	155	149	164	171	174
Number of stable contracts	117	125	112	110	111	113	113	105	126	121	116	128	121	134	151	154
Percentage of stable contracts	67.6	72.3	75.2	73.8	75.5	75.8	82.5	76.6	82.4	82.3	80.0	82.6	81.2	81.7	88.3	88.5

Similar to the second analysis, contracts can be stable from a certain percent complete but unstable from a later percent complete. Therefore, we re-examine the

stabilities and only consider the contracts stable if they remain stable from that point onward. Following the same assumptions as the second analysis about increments with no data, there are 209 contracts for each increment. Table 16 shows these adjusted results.

Table 16. Relative Interval Stability Adjusted Results of SPI(t)

SPI(t) Analysis																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Number of stable contracts	80	105	113	114	119	128	130	133	136	141	146	155	158	168	178	183
Percentage of stable contracts	38.3	50.2	54.1	54.5	56.9	61.2	62.2	63.6	65.1	67.5	69.9	74.2	75.6	80.4	85.2	87.6

Summary of SPI(t) Stability.

Table 17 provides a summary of the three analyses executed on SPI(t) stability. Bolded are the points where the percentage of stable contracts initially reaches 70, 80, and 90%. If a rule of thumb or heuristic is to be drawn from these results, the individual/organization using this heuristic must determine the relevant percentage: 70%, 80%, 90%, or some other percentage.

Table 17. Summary of SPI(t) Stability

SPI(t) Summary- Percentage of stable contracts																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Range Stability	**71.8**	78.0	**81.8**	82.8	83.3	86.1	88.0	88.5	89.5	**90.4**	91.4	92.3	94.3	95.2	96.2	97.1
Abs Interval Stability	41.1	52.2	58.4	59.8	62.2	66.5	67.9	68.9	**70.3**	72.7	75.1	78.5	79.9	**83.7**	87.1	89.0
Rel Interval Stability	38.3	50.2	54.1	54.5	56.9	61.2	62.2	63.6	65.1	67.5	69.9	**74.2**	75.6	**80.4**	85.2	87.6

Comparison Analysis

Next, we compare *categories* to see if a certain group of contracts' CPIs or SPI(t)s tends to stabilize earlier than others. To determine this, we compare the recorded ranges (between the maximum and minimum CPI or SPI(t) from specific percent complete to final) from the range definition of stability and also compare the recorded intervals (the difference between the final CPI or SPI(t) and the CPI or SPI(t) at the specific percent complete) from the absolute interval definition of stability. As stated in Chapter 3, for the range definition, we calculate the range using this formula:

$$\text{Range} = |CPImax_{X\%} - CPImin_{X\%}|$$

$CPImax_{X\%}$ equals the maximum CPI from the $X\%$ complete point to the final CPI of the contract, and $CPImin_{X\%}$ equals the minimum CPI from the $X\%$ complete point to the final CPI of the contract. We define X as a percent complete from 10 to 85 percent complete in increments of five. We use the same formula for SPI(t) by replacing CPI with SPI(t). For the absolute interval definition of stability, we calculate the interval using this formula:

$$\text{Interval} = |CPI(final) - CPI(X\%)|$$

$CPI(final)$ is the final CPI of the contract, and $CPI(X\%)$ is the CPI when the contract is $X\%$ complete. We define X as a percent complete from 10 to 85 in increments of five. We use the same formula for SPI(t) by replacing CPI with SPI(t). The relative interval definition results in the same intervals as the absolute interval (same values even though it utilizes them differently to determine stability), so they are not needed in the comparisons.

For the comparisons, we execute Kruskal-Wallis (KW) and Mann-Whitney (MW) tests to test the medians of these ranges and intervals (KW tests when there are three or more groups, MW when there are two). We cannot confidently perform t-tests, as the variables have many outliers which make the assumption of normality difficult to justify. We compare the variables (range and interval) at each percent complete increment. The hypotheses for each of our comparisons, depending on the test used, are:

KW: Ho: $\Delta_X = \Delta_Y = \Delta_Z = \cdots$ Ha: At least one median is not equal.

MW: Ho: $\Delta_X = \Delta_Y$ Ha: $\Delta_X \neq \Delta_Y$

Δ equals the median range (from the first analysis) or interval (from the second analysis) for the specific percent complete (values of 10% to 85%, in increments of five). X and Y are the two different groups in each comparison. We define what X and Y are for each test in the following sections. If a test is significant, it means one group (X or Y) has a higher median than the other. When testing ranges, whichever group possesses the higher median has a greater difference between the maximum and minimum CPI or SPI(t). When testing intervals, the group with the larger median has a greater difference between the final CPI or SPI(t) and the CPI or SPI(t) at the percent complete increment being tested.

Service.

The three services are Air Force (AF), Army, and Navy. We compare each possible pair, which results in a total of three tests (X v Y: AF v Army, Navy v Army, and AF v Navy). There are 65 AF, 45 Army, and 97 Navy contracts. We execute an overall Kruskal-Wallis test (a Mann-Whitney test that involves more than two levels) on the services to see if there is any difference in the ranges and intervals at each percent

increment overall. Then, where the overall test is significant, we perform a Mann-Whitney test. There are two contracts that are DoD contracts (not a particular service), so they are not included in the analysis. The alpha for the overall tests is 0.05, which is the family-wise error rate. Therefore, the three secondary tests performed have an alpha of 0.0167, the comparison-wise error rate (0.05 divided by 3).

CPI – Service.

To compare CPI ranges (from the range definition of stability) and intervals (from the abs interval definition of stability) by service, we execute the overall Kruskal-Wallis test to see if there is a difference between AF, Navy, and Army CPI ranges and intervals overall. Table 18 provides these results. A p-value of less than 0.05 is significant and bolded.

Table 18. Kruskal-Wallis tests on CPI Ranges and Intervals

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
P-values from overall Kruskal-Wallis tests																
CPI Ranges	**0.004**	**0.002**	**0.003**	**0.005**	**0.003**	**0.002**	**0.003**	**0.013**	**0.011**	**0.010**	**0.007**	0.058	0.347	0.264	0.096	0.170
CPI Intervals	**0.016**	0.352	0.365	0.201	0.476	**0.038**	**0.024**	0.174	0.512	0.473	**0.006**	0.354	0.574	0.583	**0.039**	0.390

Once we know which percent complete increments have significant overall tests, we execute three Mann-Whitney tests (AF v Army, Navy v Army, AF v Navy) to test if there is a difference between these specific comparisons at the significant percent complete increments. For those percent complete increments that do not have significant overall p-values, we do not execute the secondary Mann-Whitney tests because they are insignificant due to the overall test. Those cells have "-" to signify a blank cell. See Table 19 for results of the Mann-Whitney tests on CPI ranges. The significant p-values

(any p-value less than 0.0167) are bolded. For the AF v Army test, in the percent complete increments with significant p-values the AF has a smaller median than Army. For the Navy v Army test, Navy has smaller median than Army when the p-values are significant. If the p-values are not significant, we fail to reject that the medians are equal. There are no significant results in the AF v Navy test, so we fail to reject that their medians are equal at every percent complete increment.

Table 19. Mann-Whitney tests on CPI Ranges

| | P-values from Mann-Whitney tests | | | | | | | | | | | | | | | |
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	0.433	0.329	0.062	0.043	0.025	**0.013**	**0.013**	0.050	0.042	0.026	**0.011**	-	-	-	-	-
Navy v Army	**0.002**	**0.001**	**0.000**	**0.001**	**0.001**	**0.000**	**0.001**	**0.004**	**0.003**	**0.003**	**0.002**	-	-	-	-	-
AF v Navy	0.029	0.024	0.187	0.300	0.282	0.339	0.358	0.327	0.413	0.552	0.763	-	-	-	-	-

See Table 20 for the results for the comparison of the second analysis' intervals. Bolded are the significant p-values. For the significant results of the AF v Army test, the AF median is less than the Army. For the significant results of the Navy v Army test, the Navy median is less than the Army. There are no significant results in the AF v Navy test, so we fail to reject that their medians are equal.

Table 20. Mann-Whitney tests on CPI Intervals

| | P-values from Mann-Whitney tests | | | | | | | | | | | | | | | |
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	0.612	-	-	-	-	0.017	0.024	-	-	-	**0.006**	-	-	-	**0.011**	-
Navy v Army	0.019	-	-	-	-	0.040	**0.010**	-	-	-	**0.002**	-	-	-	0.080	-
AF v Navy	0.020	-	-	-	-	0.363	0.976	-	-	-	0.960	-	-	-	0.267	-

SPI(t) – Service.

To compare SPI(t) ranges and intervals by service, we execute the overall

Kruskal-Wallis test to see if there is a difference between AF, Navy, and Army SPI(t)

ranges (from the range definition of stability) and intervals (from the abs interval

definition of stability). Table 21 provides these results. A p-value of less than 0.05 is

significant and bolded.

Table 21. Kruskal-Wallis tests on SPI(t) Ranges and Intervals

P-values from overall Kruskal-Wallis tests															
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
SPI(t) Ranges	0.079	0.170	0.180	0.154	0.093	0.096	**0.048**	0.211	0.166	0.160	0.261	0.422	0.772	0.584	0.401	0.236
SPI(t) Intervals	0.087	**0.034**	0.807	0.199	0.207	**0.048**	**0.003**	0.153	0.127	0.543	0.072	0.388	0.965	0.531	0.864	0.396

Once we determine which percent complete increments have significant overall

tests, we execute three Mann-Whitney tests (AF v Army, Navy v Army, AF v Navy) to

test if there is a difference in medians between the services in each comparison at those

specific percent complete increments. For those percent complete increments that do not

have a significant overall p-value, we do not execute the secondary Mann-Whitney tests

because they are insignificant due to the overall test. Those cells have "-" to signify a

blank cell. For SPI(t) stability comparisons of the percent complete increments with

overall significant results, we first compare ranges. See Table 22 for results. There are

no significant p-values (less than 0.0167) from the tests on SPI(t) ranges. Therefore, we

fail to reject that there is no difference between the services' medians at any percent

complete increment.

Table 22. Mann-Whitney tests on SPI(t) Ranges

P-values from Mann-Whitney tests																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	-	-	-	-	-	-	0.028	-	-	-	-	-	-	-	-	-
Navy v Army	-	-	-	-	-	-	0.034	-	-	-	-	-	-	-	-	-
AF v Navy	-	-	-	-	-	-	0.466	-	-	-	-	-	-	-	-	-

See Table 23 for results of comparing the SPI(t) intervals from the absolute interval definition of stability. Bolded are the significant p-values. In the AF v Army tests, the Army medians are larger than the AF medians at the percent complete increments with significant results. In the Navy v Army tests, the Navy medians are smaller than the Army medians at the percent complete increments with significant results. At all the percent complete increments with p-values that are not significant, we fail to reject that the medians are equal.

Table 23. Mann-Whitney tests on SPI(t) Intervals

P-values from Mann-Whitney tests																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	-	**0.005**	-	-	-	0.020	**0.001**	-	-	-	-	-	-	-	-	-
Navy v Army	-	0.092	-	-	-	0.039	**0.008**	-	-	-	-	-	-	-	-	-
AF v Navy	-	0.389	-	-	-	0.562	0.370	-	-	-	-	-	-	-	-	-

Contract type.

The two types of contracts we compare are cost plus (CP) and fixed price (FP). Therefore the comparison involves only one Mann-Whitney test (CP v FP). There are 90 CP contracts and 88 FP contracts. There are 31 contracts not included in this

comparison: 16 contracts that contain both CP and FP elements and 15 Indefinite Deliver Indefinite Quantity (IDIQ) contracts. The level of significance (alpha) for this test is 0.05.

CPI - Contract Type.

Table 24 provides the results from the comparison of the ranges from the range definition of stability. There are no significant p-values, so we fail to reject that the median ranges for CP contracts and FP contracts are equal at all percent complete increments.

Table 24. Mann-Whitney tests on CPI Ranges

Percent Complete	P-values from Mann-Whitney tests

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CP v FP	0.213	0.242	0.228	0.458	0.238	0.293	0.172	0.327	0.356	0.315	0.724	0.854	0.636	0.4	0.133	0.525

Table 25 displays the comparison results of the intervals from the abs interval definition of stability. There are no significant p-values, so we fail to reject that the median intervals for CP contracts and FP contracts are equal at all percent complete increments.

Table 25. Mann-Whitney tests on CPI Intervals

Percent Complete	P-values from Mann-Whitney tests

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CP v FP	0.790	0.710	0.712	0.536	0.731	0.877	0.877	0.896	0.316	0.296	0.760	0.839	0.994	0.223	0.419	0.869

SPI(t) - Contract Type.

Using the SPI(t) ranges from the range definition of stability, Table 26 provides the results from the comparison. At every percent complete increment, the p-value is significant. For each test, the FP median is greater than the CP median.

Table 26. Mann-Whitney tests on SPI(t) Ranges

P-values from Mann-Whitney tests																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CP v FP	0.019	0.038	0.013	0.022	0.009	0.024	0.049	0.022	0.040	0.027	0.008	0.007	0.003	0.006	0.006	0.038

Table 27 displays the comparison results using the SPI(t) intervals from the abs interval definition of stability. The p-values that are significant are bolded. For each of these significant tests, the FP median is greater than the CP median. For the insignificant tests, we fail to reject that there is no difference in the medians.

Table 27. Mann-Whitney tests on SPI(t) Intervals

P-values from Mann-Whitney tests																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CP v FP	0.001	0.020	0.058	0.079	0.002	0.001	0.702	0.151	0.460	0.003	0.041	0.016	0.015	0.002	0.025	0.063

Life-cycle Phase.

The two life-cycle phases compared in this research are Development (D) and Production (P). Therefore, the comparison only involves one Mann-Whitney test (D v P). There are 102 D contracts and 102 P contracts. The level of significance (alpha) for these tests is 0.05.

CPI - Life-cycle Phase.

Table 28 displays results from testing the range values from the range definition of stability. The bolded p-values are significant. For these significant tests, the D medians are greater than the P medians. For the tests with p-values that are not significant, we fail to reject that the medians are equal.

Table 28. Mann-Whitney tests on CPI Ranges

	P-values from Mann-Whitney tests															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
D v P	**0.015**	**0.003**	**0.002**	**0.004**	**0.000**	**0.003**	**0.002**	**0.01**	**0.012**	**0.009**	**0.033**	0.079	0.202	0.333	0.315	0.198

Table 29 provides the results from the comparison of interval values from the

absolute interval definition of stability. There are no significant p-values. Therefore, we

fail to reject that the medians are equal at all percent complete increments.

Table 29. Mann-Whitney tests on CPI Intervals

	P-values from Mann-Whitney tests															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
D v P	0.470	0.274	0.385	0.483	0.198	0.904	0.154	0.814	0.704	0.207	0.511	0.870	0.296	0.809	0.649	0.132

SPI(t) - Life-cycle Phase.

See Table 30 for results from comparing ranges from the range definition of

stability. There is only one significant p-value (bolded). At this percent complete, the P

median is greater than the D median. For the tests with insignificant p-values, we fail to

reject that the D and P medians are equal.

Table 30. Mann-Whitney tests on SPI(t) Ranges

	P-values from Mann-Whitney tests															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
D v P	0.091	0.084	0.095	0.196	0.259	0.416	0.629	0.436	0.521	0.44	0.171	0.164	**0.045**	0.105	0.085	0.252

See Table 31 for results from comparing SPI(t) intervals. There is only one

significant p-value (bolded). At this percent complete, the P median is greater than the D

median. Besides that, at all the other percent complete increments, we fail to reject that

the D and P medians are equal.

Table 31. Mann-Whitney tests on SPI(t) Intervals

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
D v P	0.124	**0.036**	0.177	0.218	0.055	0.100	0.581	0.403	0.908	0.331	0.436	0.389	0.190	0.069	0.409	0.694

Platform.

For each platform, we compare its values from the range definition and abs interval definition to the values of all the others combined (for example Aircraft systems v Other platforms). The types of platforms are aircraft systems (AS), electronic/automated software systems (EAS), missile systems (MS), ordnance systems (OS), ship systems (ShS), space systems (SpS), and surface vehicle systems (SVS). To compare the median values of each platform, we first execute an overall Kruskal-Wallis test at an alpha of 0.05 to see if there is any difference in the ranges and intervals at each percent increment overall. Then, where the overall test is significant, we perform seven Mann-Whitney tests comparing each platform to the others (the tests are AS v others, EAS v others, MS v others, OS v others, ShS v others, SpS v others, and SVS v others). In total, the contracts include 59 AS, 32 EAS, 35 MS, 9 OS, 38 ShS, 13 SpS, and 10 SVS. Thirteen contracts are not included in this analysis due to data nonconformities. The overall Kruskal-Wallis test is at a 0.05 level of significance, which is the family-wise error rate. Therefore, the level of significance for each individual Mann-Whitney test is 0.0071, the comparison-wise error rate (0.05 divided by 7).

CPI – Platform.

First, we execute the overall Kruskal-Wallis test at an alpha of 0.05. See Table 32 for the results. Bolded are the significant p-values (less than 0.05). Where the p-value is

not significant, we fail to reject that the platforms have equal median ranges or intervals depending on the test.

Table 32. Kruskal-Wallis tests on CPI Ranges and Intervals

	P-values from overall Kruskal-Wallis tests															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CPI Ranges	**0.036**	0.089	0.400	0.441	0.220	0.200	0.495	0.364	0.369	0.463	0.692	0.511	0.654	0.775	0.878	0.170
CPI Intervals	0.182	**0.013**	0.202	0.938	0.152	0.077	0.340	0.161	0.077	0.313	0.805	0.918	0.630	0.470	0.809	0.222

Table 33 provides the results of the Mann-Whitney tests which compare the ranges from the range definition of stability at the percent complete increments that generate significant p-values from the overall Kruskal-Wallis tests. For those percent complete increments that do not have a significant overall p-value, we do not execute the secondary Mann-Whitney tests because they are insignificant due to the overall test. Those cells have "-" to signify a blank cell. From the Mann-Whitney test, there are no significant p-values (less than 0.0071). Therefore, we fail to reject that the platforms' CPI ranges have equal medians.

Table 33. Mann-Whitney tests on CPI Ranges

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
	P-values from Mann-Whitney tests															
AS v Others	0.044	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EAS v Others	0.069	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MS v Others	0.037	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OS v Others	0.462	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ShS v Others	0.040	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SpS v Others	0.675	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SVS v Others	0.546	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 34 contains the results of comparing CPI intervals from the abs interval definition of stability. The significant p-values are bolded. For the EAS v Others significant test at 15 percent complete, the EAS median is greater than the median of the other platforms. For the MS v Others significant test at 15 percent complete, the MS median is less than the median of the other platforms. For all tests with p-values that are not significant, we fail to reject that the medians are equal.

Table 34. Mann-Whitney tests on CPI Intervals

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
P-values from Mann-Whitney tests																
AS v Others	-	0.453	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EAS v Others	-	**0.002**	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MS v Others	-	**0.006**	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OS v Others	-	0.769	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ShS v Others	-	0.367	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SpS v Others	-	0.487	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SVS v Others	-	0.586	-	-	-	-	-	-	-	-	-	-	-	-	-	-

SPI(t) – Platform.

We execute the overall Kruskal-Wallis test at an alpha of 0.05. See Table 35 for the results. There are no significant p-values. Therefore, we fail to reject that the platforms have equal median ranges or intervals depending on the test. Since there are none with significant results in the overall tests, we do not execute the Mann-Whitney tests.

Table 35. Kruskal-Wallis tests on SPI(t) Ranges and Intervals

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
P-values from overall Kruskal-Wallis tests																
SPI(t) Ranges	0.689	0.806	0.841	0.845	0.854	0.909	0.887	0.896	0.859	0.876	0.917	0.840	0.993	0.886	0.870	0.269
SPI(t) Intervals	0.094	0.286	0.446	0.512	0.849	0.834	0.220	0.477	0.928	0.723	0.712	0.882	0.680	0.956	0.894	0.793

Conclusion

In Chapter 4, we provide the results from the variance analysis of CPI and SPI(t) among 209 contracts and comparison analysis of the CPI and SPI(t) stabilities by service, contract type, life-cycle phase, and platform. Chapter 5 utilizes these results to draw conclusions on the characteristics of CPI and SPI(t) stabilities in attempt to answer our research questions stated in Chapter 1.

V. Conclusion

Introduction

This thesis examines CPI and SPI(t) stability in 209 DoD ACAT 1 program contracts over the past 25 years. The CPI helps estimate the final cost of a program. If the efficiency index stabilizes, or does not change more than a certain amount, program managers can be more confident in the forecasted final cost/schedule of the program. Stability findings also have desirable properties for analysis of TCPI. Given stability, if the TCPI is much higher than the current CPI, CPI stability tells us that the project is unlikely to reach that TCPI and will most likely finish over budget. Therefore, CPI stability could help make more informed decisions on programs and ultimately save money by making smarter decisions to continue or stop funding certain programs. SPI(t) could be used with the EVM measurements of CPI and SPI, and therefore it could exhibit similar stability characteristics and provide similar benefits.

Stability History

In 1993, Scott Heise and David Christensen conducted the landmark study on CPI stability. After analyzing 155 DoD contracts, they discovered that the cumulative CPI was stable from when the contract was 20 percent complete for 86 percent of the contracts (Christensen and Heise, 1993: 5). They defined stability as when the maximum and minimum CPIs, from a specific percent complete to the end of the contract, has a difference of less than or equal to 0.20. In the final of a series of research projects, Christensen and Carl Templin performed a study on CPI stability in 2002, looking at 240 DoD contracts. They re-defined stability as when the difference between the final CPI and the CPI at 20% complete is greater than or equal to 0.10 in order to test the current

rule of thumb that morphed from Heise's work that states, "the cumulative [CPI] will not change by more than 0.10 from its value at the 20 percent completion point" (Christensen and Templin, 2002: 1). They found that the CPI did not change by more than 0.10, "with only few exceptions" (Christensen and Templin, 2002: 8). In 2008, Kym Henderson and Ofer Zwikael examined CPI stability in contracts outside the DoD. Analyzing a dataset of 45 projects from the United Kingdom, Israel, and Australia, they found that the CPI did not stabilize by 20 percent complete, but "often later than 80 percent complete" (2008: 9). For their study, they used the same definition of stability as Christensen and Templin's 2002 study. While these past studies provide different conclusions about CPI stability, until now there has been no research on SPI(t) stability in DoD contracts.

Research Questions Answered

This research focuses on both CPI and SPI(t) stability properties in DoD ACAT I programs. Using the results from Chapter 4, the six research questions from Chapter 1 can be answered. Results and discussion of each of these questions follows.

1. **When in a program's life does the CPI tend to stabilize?**

Table 36 displays the results of the CPI stability research. It provides the percentage of contracts that possess stable CPIs at specific percent complete points for all *three definitions* of stability utilized throughout this research.

Table 36. CPI Stability Percentages

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CPI Summary- Percentage of stable contracts																
Range Stability	**72.7**	78.5	**82.8**	83.7	87.6	**91.4**	93.3	94.3	95.2	96.2	98.1	98.6	98.6	99.0	99.5	99.5
Abs Interval Stability	39.7	50.7	56.0	59.8	61.7	66.5	69.9	**72.2**	76.1	**80.9**	83.7	85.6	89.0	**91.9**	95.7	98.6
Rel Interval Stability	39.7	50.7	55.0	57.9	60.3	64.1	67.5	69.4	**72.7**	77.5	**81.8**	83.7	86.6	**90.4**	94.3	97.6

The CPI's stability tendencies depend upon the definition of stability. The range definition of stability identifies stability as when the difference between the maximum and minimum CPI or SPI(t) between a specific percent complete and the final point is less than 0.2. The absolute interval definition identifies stability as when the final CPI (or SPI(t)) is within 0.10 of the CPI (or SPI(t)) at a specific percent complete. The relative interval definition of stability, which is also referred to in the literature, is when the difference between the final CPI and the CPI of a specific percent complete is less than or equal to plus or minus 10% of the CPI at the specific percent complete.

Using the range stability definition, 72.7 percent of the contracts have stable CPIs from 10 percent complete, 82.8 percent are stable from 20 percent complete, and 91.4 percent from 35 percent complete. Using the absolute interval definition of stability, 72.2 percent of the contracts possess stable CPIs from 45 percent complete, 80.9 percent are stable from 55 percent complete, and 91.9 percent stable from 75 percent complete. With the relative interval stability definition, 72.7 percent of contracts have stable CPIs from 50 percent complete, 81.8 percent stable from 60 percent complete, and 90.4 percent from 75 percent complete. These values are bolded in the table to provide when the

stable percentage reaches 70, 80, and 90 percent. These stability percentages (70, 80, and 90) are highlighted because the literature review of Chapter 2 indicates they may be important milestones for determining the stability tendencies. For example, from these numbers, one can state at least 90 percent of the contracts have stable CPIs from 35 percent complete, using the range stability definition.

Depending on the definition of stability, the results in Table 36 *both support and contradict* the "stability rule" within the DoD and the findings of earlier research. With the range definition of stability, 82.8 percent of the contracts possess a stable CPI when the contract is 20 percent complete, which is similar to the 86 percent of contracts being stable at the 20% completion point from Christensen and Heise's research. However, with either of the interval stability definitions, only about 55 percent of the contracts possess stable CPIs at 20 percent complete. This contradicts the interpretation of the stability rule that states the "cumulative CPI will not change by more than 0.1 from its value at the 20 percent completion point" (Christensen and Templin, 2002: 5).

2. When in a program's life does the SPI(t) tend to stabilize?

Table 37 displays the results of the SPI(t) stability research. It contains the percentage of contracts that possess a stable SPI(t) at a specific percent complete for all three definitions of stability.

Table 37. SPI(t) Stability Percentages

SPI(t) Summary- Percentage of stable contracts															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Range Stability	**71.8**	78.0	**81.8**	82.8	83.3	86.1	88.0	88.5	89.5	**90.4**	91.4	92.3	94.3	95.2	96.2	97.1
Abs Interval Stability	41.1	52.2	58.4	59.8	62.2	66.5	67.9	68.9	**70.3**	72.7	75.1	78.5	79.9	**83.7**	87.1	89.0
Rel Interval Stability	38.3	50.2	54.1	54.5	56.9	61.2	62.2	63.6	65.1	67.5	69.9	**74.2**	75.6	**80.4**	85.2	87.6

As with the CPI, the SPI(t)'s stability behaves differently depending on the stability definition used. With the range stability definition, 71.8 percent of contracts possess stable SPI(t)s from 10 percent complete, 81.8 percent are stable from 20 percent complete, and 90.4 percent stable from 55 percent complete. With the absolute interval definition, 70.3 percent of the contracts have stable SPI(t)s from 50 percent complete, and 83.7 percent are stable from 75 percent complete. With the relative interval definition, 74.2 percent of the contracts contain stable SPI(t)s from 65 percent complete, and 80.4 percent are stable from 75 percent complete. For both interval definitions, stability never reaches 90 percent of the contracts until after the 85 percent complete point.

Utilizing the range definition, SPI(t) stability is found to be very similar to CPI stability. The 80 percent contract stability threshold is reached for both CPI and SPI(t) stability at the 20 percent completion point. In contrast, using either interval definition, the SPI(t) is not stable until much later than the corresponding CPI. The SPI(t) interval definition obtains stability for 80 percent of the contracts once the contracts are 70 percent complete, while the CPI interval definitions were much earlier at 55 and 60 percent respectively.

3. **What differences in CPI and SPI(t) stabilities exist between branches of the DoD?**

To determine if CPIs and SPI(t)s have different stabilities between the United States Air Force (AF), Navy, and Army, we compare their median ranges and intervals. "Range" refers to the calculation from the range definition of stability:

$$\text{Range} = |CPImax_{X\%} - CPImin_{X\%}|$$

$CPImax_{X\%}$ equals the maximum CPI from the $X\%$ complete point to the final CPI of the contract, and $CPImin_{X\%}$ equals the minimum CPI from the $X\%$ complete point to the final CPI of the contract. "Interval" refers to the difference between the final CPI or SPI(t) and the CPI or SPI(t) at X percent complete (this interval calculation is the same calculation from both interval definitions of stability). For both these formulas, X is a percent complete from 10 to 85 in increments of five. The same formulas determine SPI(t)'s ranges and intervals by replacing CPI with SPI(t). An overall Kruskal-Wallis test with an alpha of 0.05 determines if all services are different. Then, only when the overall Kruskal-Wallis has significant results, we utilize Mann-Whitney tests to compare pairs of services' median range and interval (AF v Navy, AF v Army, Navy v Army). If a Service has a smaller median CPI or SPI(t) range or interval than the Service it is compared to, it's CPI or SPI(t) is more stable when using that particular stability definition.

For all the tests that compare AF and Navy CPIs and SPI(t)s, there is no statistical difference. Therefore, there is no difference between the CPI and SPI(t) stabilities of AF and Navy contracts, as calculated by the ranges in the range definition of stability and the

intervals in the absolute interval definition of stability. Because of this, these results are not in the tables of this section.

When examining the CPI ranges calculated in the range definition of stability, there are significant results when comparing AF with Army and Navy with Army. Table 38 summarizes these results. Cells with "-" signify no significant results, so there is no statistical difference in the CPI ranges between the two services compared in that cell.

Table 38. Comparing CPI Ranges by Service

Percent Complete	P-values from Mann-Whitney tests on CPI Ranges															
	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	-	-	-	-	-	0.013	0.013	-	-	-	0.011	-	-	-	-	-
Navy v Army	0.002	0.001	0.000	0.001	0.001	0.000	0.001	0.004	0.003	0.003	0.002	-	-	-	-	-

For each increment from 10 percent complete to 60 percent complete, Navy has a smaller median CPI range (range in terms of the calculated ranges from the range stability definition) than Army. This means the Navy contracts' CPIs did not change as much as the Army's. There are only three significant tests when comparing AF with Army, which all state that AF contracts' CPI ranges have a smaller median than Army's.

Comparing CPI intervals calculated from the absolute interval definition of stability, there are only four tests with significant results (see Table 39). Between AF and Army CPI intervals, AF has a smaller median in the two tests that are statistically significant. For the results from comparing Navy and Army, Navy has a smaller median. All other cells have "-" to signify that there is no significant difference between the two services compared. Therefore, Army either has statistically the same median or a greater median in CPI intervals than the other two services. This is evidence that AF and Navy

contracts' CPIs are more stable than Army's, but not necessarily all the time. We can definitely state that Army contracts' CPIs, however, are not more stable.

Table 39. Comparing CPI Intervals by Service

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	-	-	-	-	-	-	-	-	-	-	0.006	-	-	-	0.011	-
Navy v Army	-	-	-	-	-	-	0.010	-	-	-	0.002	-	-	-	-	-

There are no significant results when comparing SPI(t) ranges between services. Therefore, they exhibit no differences in SPI(t) stability using the range definition. Comparing the intervals from the SPI(t) interval stability definition, there are three significant results, as shown in Table 40. In these three tests, the Army median is higher than the respective other (AF or Navy depending on the test). Therefore, Army contracts do not have more stable SPI(t)s than Navy or AF when using either interval definition of stability.

Table 40. Comparing SPI(t) Intervals by Service

Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF v Army	-	0.005	-	-	-	-	0.001	-	-	-	-	-	-	-	-	-
Navy v Army	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-	-

In summary, there is no difference in AF and Navy for all the tests undertaken. Additionally, there is no difference in all three services using SPI(t) ranges. However, Navy's CPI ranges have smaller medians than Army from 10 to 60 percent complete. The AF v Army tests only have a few significant differences, where AF has smaller

medians when using CPI ranges, CPI intervals, and SPI(t) intervals. This is consistent

with earlier research findings by Christensen and Templin in 2002. Using the absolute

interval definition, they found that from 20% complete Army's mean interval was larger

than AF and Navy's mean intervals (Christensen and Templin, 2002: 15). Although

Christensen and Templin tested the *mean* value, and this research tested the *median*, their

data displays similar results that Army's CPI tends to change more than AF or Navy's.

4. What differences in CPI and SPI(t) stabilities exist between contract types?

The next comparison is of median ranges and intervals (ranges and intervals

defined in the third research question section) by contract type. The two categories

compared are Cost Plus (CP) and Fixed Price (FP). Mann-Whitney tests at an alpha of

0.05 compare the medians at every percent complete (in increments of five) from 10% to

85%.

There are no significant results from the Mann-Whitney tests on CPIs of CP and

FP contracts by either stability definition. Therefore, there is no difference in CPI

stability between the two contract types. This finding is consistent with past research.

Using the absolute interval definition of stability, Christensen and Templin found the

mean intervals of CP contracts to be very similar to FP contracts (2002:15). Christensen

and Heise found only a slight difference (of just 1 percent) in the mean stability points of

CP and FP contracts, using the range definition (1993: 10).

We do observe differences in SPI(t) stabilities between contract types. Table 41

displays the significant results. For these significant tests, using both range and interval

definitions of stability, FP contracts have a greater median than CP. When using the

range definition of stability, we can conclude that SPI(t)s of CP contracts tend to be more

74

stable than of FP contracts. Using the interval definitions of stability, there are not

significant differences all the way through the contract's life. However, there is still

strong evidence that shows the SPI(t)s of CP contracts tend to be more stable than of FP

contracts since the majority of the tests demonstrated this.

Table 41. Comparing SPI(t) Stability by Contract Type

P-values from Mann-Whitney tests on SPI(t) - CP v FP															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Ranges	0.019	0.038	0.013	0.022	0.009	0.024	0.049	0.022	0.040	0.027	0.008	0.007	0.003	0.006	0.006	0.038
Intervals	0.001	0.020	-	-	0.002	0.001	-	-	-	0.003	0.041	0.016	0.015	0.002	0.025	-

These SPI(t) stability results are surprising because CP contracts are typically

utilized when there is more uncertainty involved in the contract. A typical example of a

CP contract would be the development effort for a new bomber. This type of contract

would logically be thought to have more variation in schedule performance, but the

results found here are contrary.   One possible explanation is that the contractors may add

contract change proposals or an engineering change to an FP program when the

contractor is losing money. By attempting to increase the scope and receive more money,

the schedule suffers, as there will be no money for overtime or to hire more personnel in

an attempt to catch up. Also, contractors may use "other techniques [including]

negotiating meaninglessly general statements of work, or agreeing to successive, after-

the-fact, incremental fixed-price contracts that simply reimburse contractors for work

already performed" which will ultimately worsen the performance of FP contracts (CBO,

1982: 11). This explanation sheds some light on these findings but is not conclusive.

This is an area for future research.

5. **What differences in CPI and SPI(t) stabilities exist between life-cycle phases?**

Mann-Whitney tests (with alpha of 0.05) determine if differences exist in CPI and SPI(t) median ranges and intervals between the two main life-cycle phases, Production (P) and Development (D).

When comparing CPI stabilities, there are only significant results when using the range definition of stability (absolute interval definition yielded no significant results). See Table 42 for these results. From 10 percent complete to 60 percent complete, the calculated CPI ranges from contracts in D phase have a higher median than CPI ranges from contracts in P phase. Therefore, we can conclude P contracts tend to have more stable CPIs than D contracts up until about the 62.5 percent complete point. This finding that P contracts have more stable CPIs is logical because a P contract is typically an iterative process of recreating something that has already been developed, whereas a D contract is creating something new. This finding is consistent with past research. Christensen and Heise found the mean stability point, using the range definition, of P contracts to be earlier than D contracts. They found the mean stability points to be 9 percent complete for P and 15 percent complete for D (Christensen and Heise, 1993: 10).

Table 42. Comparing CPI Ranges by Life-cycle Phase

P-values from Mann-Whitney tests on CPI – P v D															
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Ranges	0.015	0.003	0.002	0.004	0.000	0.003	0.002	0.01	0.012	0.009	0.033	-	-	-	-	-

When comparing SPI(t) stability using either definition (range or interval), there were two tests that yielded significant results (one from each definition, see Table 43). For both of these tests, the SPI(t) of P contracts have a greater median range or interval

than D contracts. Since there is only one significant test for each definition, it is difficult to conclude that there is any difference in SPI(t) stabilities between life-cycle phases.

Table 43. Comparing SPI(t) by Life-cycle Phase

P-values from Mann-Whitney tests on SPI(t) – P v D																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Ranges	-	-	-	-	-	-	-	-	-	-	-	-	0.045	-	-	-
Intervals	-	0.036	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Therefore, the results show that the P contracts are more stable than D contracts in terms of CPI, using the range definition of stability. From testing CPI intervals, SPI(t) ranges, and SPI(t) intervals, there is little to zero statistically significant evidence of any differences in stability.

6. **What differences in CPI and SPI(t) stabilities exist between different military platforms?**

To compare military platforms, the Kruskal-Wallis test with an alpha of 0.05 determines if there is a difference between the seven categories of platforms: aircraft systems (AS), electronic/automated software systems (EAS), missile systems (MS), ordnance systems (OS), ship systems (ShS), space systems (SpS), and surface vehicle systems (SVS). Where results are statistically significant, Mann-Whitney tests with alphas of 0.0071 conducted on each category against the others combined determine if individual platforms have different stabilities than the rest. We execute tests at each percent complete from 10 to 85 in increments of 5.

There are no significant results when comparing CPI ranges and SPI(t) ranges and intervals. Therefore, no statistically significant differences exist in CPI ranges and SPI(t)

ranges and intervals between the different platforms. There are only two significant

results from comparing CPI intervals: EAS v Others (with a p-value 0.002) where EAS

has a greater median, MS v Others (with a p-value 0.006) where MS has a smaller

median. Since there is only one test out of sixteen (one for each 5 percent complete

increment) that shows significant results for the EAS platform against all other platforms

or MS against all other platforms, it is difficult to conclude that these platforms have

different stability characteristics than the rest.

Limitations

There are several limitations in this research. First, Over-Target-Baselines

(OTBs) contracts removed from the dataset total to about 20% of all the available

contracts. While removal of OTBs is consistent with previous DoD stability research,

this is a large amount of data that was not able to be utilized. Second, the dataset

contains contracts from only DoD ACAT I programs. The results should not be

generalized to programs outside the DoD without confirmatory research being conducted

within a program portfolio of interest and may not even be properly generalized to non-

ACAT I programs in the DoD. Thirdly, as described in Chapter 2, the EVM Central

Repository contains data from contracts that followed different sets of reporting

requirements due to changes in EVM policy over time. Therefore, the quality of EVM

data may have been affected.

Further Discussion

Some researchers and authors have taken the original Christensen and Heise

research (1993) and indicated that their results were "generalizable," though the authors

never made this claim. The concept of generalizability, or a rule of thumb, is an

empirical matter. No claim is made in this paper either way. It is left to those EVM

practitioners in the field to determine the applicability of the stability findings in this

paper to analysis of their program(s). Regardless of their determination, the results of

this analysis are only directly applicable to DoD ACAT I programs which have not

undergone an OTB. Importantly, according to Kristine Thickstun, there is no way to

predict if a contract will have an OTB (Thickstun, 2010). Therefore, although this

research shows certain stability characteristics when the contract *does not* have an OTB,

the question of whether the contract will have an OTB or not remains. Thus, the

applicability of stability properties to the contract remains tied to the unresolved question

of being able to predict whether the contract will be OTB or not. This limitation is true

of all the past DoD research as well since they too removed OTBs from the analysis.

Conclusion

The definition of "stability" has (understandably) morphed over time. To answer

the question of stability, then, is intricately tied to the definition used. This research finds

that CPI stability utilizing the "range" definition behave similar to past research and the

"stability rule" but CPI stability utilizing the interval definition stabilize later than the

original "stability rule" states. SPI(t) behaves very similar to CPI when using the range

definition of stability but stabilizes later in a contracts life using the interval definitions.

From comparing CPI and SPI(t) stabilities among services, AF and Navy have very

similar stabilities for both indices, and Army contracts' CPIs and SPI(t)s are either the

same or less stable than AF and Navy's. When comparing contract types, we observe

that CPI behaves the same for CP and FP contracts, but the SPI(t) tends to be more stable

in CP contracts. Between life-cycle phases, SPI(t) stability is very similar, but production

contracts are more stable than development contracts in terms of CPI, using the range definition of stability (CPI intervals have no difference). Comparisons between platforms show that the different platforms have no difference in CPI and SPI(t) stabilities.

The different definitions of stability have their advantages and disadvantages. The range definition is less dependent on a specific percent complete since it uses the maximums and minimums, whereas the interval definitions are more reliant on a single percent complete since that single point determines stability or not. The range definition takes into account the entire contract's life after a specific percent complete, but the absolute interval looks at a single point and compares it to the final. The range definition, however, is a little more complicated to comprehend and apply especially when using to prospectively predict contract performance. You do not know if the current CPI is the maximum, minimum, or somewhere in the middle of the contract's entire performance, so it is difficult to use a definition that utilizes the maximum and minimum. The interval definitions are easier to apply and more conservative, ultimately predicting a range of plus or minus 0.10 around the given CPI. The relative interval definition is simply another interpretation of the absolute interval definition. Therefore, if we have to choose a single definition for stability to use in the future, we recommend the absolute interval definition. It is more dependent on a single percent complete, but it is easier to understand and more conservative. These two characteristics are important to program offices as they examine the performance of contracts.

Further Research

The comparison analysis results determined that Cost Plus contracts have more stable SPI(t)s than Fixed Price contracts. Possible explanations include the contractors

performing specific tactics to reimburse or protect themselves that cause the schedule performance to suffer. These explanations shed some light on these findings but are not conclusive. This is an area for future research.

For this research, we studied DoD contracts but had to remove any contract with an OTB from the analysis. If there is a way to include the OTBs in stability research, it would capture more of the available data that could then be utilized for research as well as improving our understanding of the performance characteristics of OTB contracts.

Some of the contracts were not stable until near the end of the contract. Are there improved ways to accurately predict if a contract will have a stable CPI or SPI(t)? Do contracts with stable CPIs or SPI(t)s have common characteristics, outside the characteristics we tested? Do the contracts that have unstable CPIs or SPI(t)s have common characteristics? As noted in past research (Henderson and Zwikael, 2008), it may be helpful to recognize more contract characteristics that improve or worsen CPI or SPI(t) stability. There may even be characteristics beyond what is displayed in the contract performance reports that explain SPI(t) and CPI behavior. Some possible characteristics are listed here: amount of overtime for the reporting period, drawing releases versus their schedule, amount of weight growth, number of people leaving the program, changes in overhead rates.

When analyzing stability using either interval definition, we notice some contracts' CPIs and SPI(t)s change from being stable to unstable and then back to being stable. After being within 0.1 of the final CPI or SPI(t), there is a time window when the contracts' CPIs or SPI(t)s are not within 0.1 of the final CPI or SPI(t). What causes them to change from being stable to unstable, and then what causes them to go back to being

stable? Is there an event or performance characteristics common among these contracts

that causes this? This research could help further the benefits of the long history of

stability research.

Appendix A: EVM Criteria

The EVMS criteria, published in American National Standards Institute/Electronic
Industries Alliance (ANSI/EIA) Standard 748, Earned Value Management
Systems, are the following:

Criterion 1. Define the authorized work elements for the agency. A WBS, tailored for
effective internal management control, is commonly used in this process.

Criterion 2. Identify the organizational structure including the major contractors
responsible for accomplishing the authorized work, and define the organizational
elements in which work will be planned and controlled.

Criterion 3. Provide for the integration of the agency's planning, scheduling, budgeting,
work authorization and cost accumulation processes with each other, the WBS,
and the OBS.

Criterion 4. Identify the organization or function responsible for controlling overhead
(indirect costs).

Criterion 5. Provide for integration of the WBS and the organizational structure in a
manner that permits cost and schedule performance measurement by elements of
either or both structures as needed.

Criterion 6. Schedule the authorized work in a manner that describes the sequence of
work and identifies significant task interdependencies required to meet all
authorized requirements

Criterion 7. Identify physical products, milestones, technical performance goals, or other
indicators that will be used to measure progress.

Criterion 8. Establish and maintain a time-phased budget baseline, at the control account

   level, against which performance can be measured. Budget for far-term efforts

   may be held in higher-level accounts until an appropriate time for allocation at the

   control account level. Initial budgets established for performance measurement

   will be based on either internal management goals or the external customer-

   negotiated target cost including estimates for authorized but undefined work.

Criterion 9. Establish budgets for authorized work with identification of significant cost

   elements (labor, material, etc.) as needed for internal management and for control

   of contractors.

Criterion 10. To the extent it is practical to identify the authorized work in discrete work

   packages, establish budgets for this work in terms of dollars, hours, or other

   measurable units. Where the entire control account is not subdivided into work

   packages, identify the far term effort in larger planning packages for budget and

   scheduling purposes.

Criterion 11. Provide that the sum of all work package budgets plus planning package

   budgets within a control account equals the control account budget.

Criterion 12. Identify and control level of effort activity by time-phased budgets

   established for this purpose. Only that effort which is immeasurable or for which

   measurement is impractical may be classified as level of effort.

Criterion 13. Establish overhead budgets for each significant organizational component

   of the company for expenses that will become indirect costs. Reflect in the

   budgets, at the appropriate level, the amounts in overhead pools that are planned

   to be allocated as indirect costs.

Criterion 14. Identify management reserves and undistributed budget.

Criterion 15. Provide that the allocated budget is reconciled with the sum of all internal
budgets and management reserves.

Criterion 16. Record direct costs in a manner consistent with the budgets in a formal
system controlled by the general books of account.

Criterion 17. Summarize direct costs from control accounts into the WBS without
allocation of a single control account to two or more WBS elements.

Criterion 18. Summarize direct costs from the control accounts into the agency's
organizational elements without allocation of a single control account to two or
more organizational elements.

Criterion 19. Record all indirect costs that will be allocated to the agency.

Criterion 20. Identify unit costs, equivalent units costs, or lot costs when needed.

Criterion 21. For EVMS, the material accounting system will provide for:

•Accurate cost accumulation and assignment of costs to control accounts
in a manner consistent with the budgets using recognized, acceptable, costing
techniques.

•Cost performance measurement at the point in time most suitable for the
category of material involved, but no earlier than the time of progress payments or
actual receipt of material.

•Full accountability of all material purchased including the residual
inventory.

Criterion 22. At least on a monthly basis, generate the following information at the

control account and other levels as necessary for management control using actual

cost data from, or reconcilable with, the accounting system:

•Comparison of the amount of planned budget and the amount of budget

earned for work accomplished. This comparison provides the schedule variance.

•Comparison of the amount of the budget earned with the actual (applied

where appropriate) direct costs for the same work. This comparison provides the

cost variance.

Criterion 23. Identify, at least monthly, the EV Variance between both planned and actual

schedule performance and planned and actual cost performance, and provide the

reasons for the variances in the detail needed by management.

Criterion 24. Identify budgeted and applied (or actual) indirect costs at the level and

frequency needed by management for effective control, along with the reasons for

any significant variances.

Criterion 25. Summarize the data elements and associated variances through the

organization and/or WBS to support management needs and any customer

reporting specified in the contract.

Criterion 26. Implement managerial actions taken as the result of earned value

Criterion 27. Develop revised EAC based on performance to date, commitment values for

material, and estimates of future conditions. Compare this information with the

performance measurement baseline to identify variances at completion important

to company management and any applicable customer reporting requirements

including statements of funding requirements.

Criterion 28. Incorporate authorized changes in a timely manner, recording the effects of such changes in budgets and schedules. In the directed effort prior to negotiation of a change, base such revisions on the amount estimated and budgeted to the organizations.

Criterion 29. Reconcile current budgets to prior budgets in terms of changes to the authorized work and internal replanning in the detail needed by management for effective control.

Criterion 30. Control retroactive changes to records pertaining to work performed that would change previously reported amounts for actual costs, earned value, or budgets. Adjustments should be made only for correction of errors, routine accounting adjustments, effects of customer or management directed changes, or to improve the baseline integrity and accuracy of performance measurement data.

Criterion 31. Prevent revisions to the agency budget except for authorized changes.

Criterion 32. Document changes to the performance measurement baseline.

Appendix B: List of Programs in the Dataset

The programs in the dataset include the following:

5D-3 Weather Satellite

AAAV (Advanced Amphibious Assault Vehicle)

AEGIS Combat System (ACS)

AFATDS (Advanced Field Artillery Tactical Data System)

AH-64 Longbow Apache

AIM-54C Phoenix Missile

AIM-9X Sidewinder

ALQ-135 Jammer

AMDR - Air & Missile Defense Radar

AMRAAM Pre-Planned Product Improvement (P3I)

AN/BSY-2(v) Submarine Combat Systems Program

AN/FPS-118 Over-The-Horizon RADAR

AOE class ship

ASPJ (Airborne Self-Protection Jammer)

ATACMS (Army Tactical Missile System)

ATACMS- BAT- Brilliant Anti-Armor Technology

ATIRM/CMWS (Advanced Threat Infrared Countermeasure/Common Missile Warning
System)

AWACS-RSIP (Airborne Warning and Control System - Radar System Improvement
Program)

B-2 EHF SATCOM AND COMPUTER INCREMENT I and II– B-2 Advanced

Extremely High Frequency SatCom Capability

BLACK HAWK UPGRADE (UH-60M) - Utility Helicopter Upgrade Program

C-17A

C-5 RERP - C-5 Aircraft Reliability Enhancement and Re-engining Program

CBU-97A/B Sensor Fused Weapons (Cluster Bomb Unit)

CEC – Cooperative Engagement Capability

CG class ships

Chem Demil - CMA (Anniston Chemical Demilitarization Facility)

Chem Demil - CMA (Pine Bluff Chemical Agent Disposal Facility)

Chem Demil-Newport (Newport Chemical Agent Disposal Facility)

Combat Service Support Control System (CSSCS)

Continuation of ops, repair, sustainment support for GPS Block II/IIA satellite program

Conventional Sealift Prepositioned/Surge ship

CVN-76 Supercarrier

DDG 1000 - ZUMWALT CLASS Destroyer

DDG 51- ARLEIGH BURKE CLASS Guided Missile Destroyer

DDG class Ship

E-2D AHE - E-2D Advanced Hawkeye

EA-18G - Airborne Electronic Attack variant of the F/A-18 aircraft

EXCALIBUR - Family of Precision, 155mm Projectiles

F/A-18E/F

F/A-18E/F - SUPER HORNET Naval Strike Fighter

F-22 - RAPTOR Advanced Tactical Fighter

FAAD C21 BLK II-IV (Forward Area Air Defense Command, Control, and Intelligence)

FCS - Future Combat Systems

FDS UWS (Fixed Distributed System Underwater Segment)

GMLRS (Guided Multiple Launch Rocket System)

GMLRS/GMLRS AW - Guided Multiple Launch Rocket System/Guided Multiple

> Launch Rocket System Advanced Warhead

GPS OCX – Global Positioning Satellite Next Generation Control Segment

GPS Phase III User Equipment

Granite Sentry Operational Acceptance System

HC/MC-130 Recapitalization

JAGM – Joint Air-to-Ground Missile

JASSM (JASSM/JASSM-ER) – Joint Air-to-Surface Standoff Missile

JATAS (Joint and Allied Threat Awareness System)

JAVELIN LRIP (Anti-tank missile)

JDAM (Joint Direct Attack Munition)

JHSV - Joint High Speed Vessel

JLTV – Joint Lightweight Tactical Vehicle

JPALS - Joint Precision Approach and Landing System

JPATS (Joint Primary Aircraft Training System)

JSOW (Joint Standoff Weapon)

JTN - Joint Tactical Networks

LCAC (Landing Craft Air Cushion)

LCS - Littoral Combat Ship

LHD class Ship

MH-60R - Multi-Mission Helicopter Upgrade

MHC (Mine Hunter Coastal)

MIDS-LVT (Multifunctional Information Distribution System- Low Volume Terminal)

Milstar (Military Strategic and Tactical Relay)

Minuteman III GRP (Guidance Replacement Program)

MK 50 Torpedo

MPS – Mission Planning System

MQ-1C GRAY EAGLE

MQ-4C Triton (Formerly BAMS)

MQ-9 Reaper

MV-22

NMT (Navy Multiband Terminal)

PAC-3 - Patriot Advanced Capability 3

Patriot missile

PATRIOT/MEADS CAP - Patriot/Medium Extended Air Defense System Combined

      Aggregate Program

Peacekeeper Intercontinental Ballistic Missile- Missilie guidance and control system

Peacekeeper Post Boost Vehicle (PBV)

Phalanx Close-In Weapon System (CIWS)

Phase II of Service Lift Extension Program for Defense Satellite Communication System

Rocket Motor and a ShortenedControl Auction System into the AIM-120C Advanced

Medium Range Air-to-AirMissiles (AMRAAM)

SADARM (Sense And Destroy Armor)

Satellite Readout Station Upgrade(SRSU) antennas

SH-60R

Space Defense Operations Center (SPADOC)

Special Operations Forces Air Training System (SOF ATS) for the MC-130H&E aircraft

SSBM-741 Submarine

SSN-688 ATTACK SUB

STD MSL (Standard Missile) 2 (BLKS I-IV)

STINGER RMP (Reprogrammable Microprocessor)

Strategic Sealift Ship

T800 Helicopter Engine

THAAD GBR (Theater High Altitude Area Defense, Ground-Based Radar)

Threaded Fastener

Trident II Missile

Trident Missile Program

UH-47 Chinook

USS NORTH CAROLINA (SSN 777) Post Shakedown Availability (PSA)

V-22 - OSPREY Joint Advanced Vertical Lift Aircraft

VTUAV - Vertical Takeoff and Land Tactical Unmanned Air Vehicle (Fire Scout)

WGS – Wideband Global SATCOM Program

WIN-T Inc. 2 - Warfighter Information Network Tactical Increment 2

Bibliography

Abba, Wayne. "The Trouble with Earned Schedule," *The Measurable News*, Fall, Issue 4: 28-30 (2008).

Bowman, Thomas L. Personal Interview, 5 September 2013.

Christensen, David S. "Project Advocacy and the Estimate At Completion Problem," *Journal of Cost Analysis*, Spring 1996: 35-60 (1996).

-----. "Using the Earned Value Cost Management Report to Evaluate the Contractor's Estimate at Completion." *Acquisition Review Quarterly*, 19 (Summer):283-296 (1999).

Christensen, David S. and Kirk I. Payne. "Cost Performance Index Stability- Fact or Fiction?" *Journal of Parametrics*, 10:27-40 (1992).

Christensen, David S. and Scott R. Heise "Cost Performance Index Stability," *National Contract Management Journal*, 25:7-15 (1993).

Christensen, David S. and Carl Templin. "EAC Evaluation Methods: Do They Still Work?" *Acquisition Review Quarterly*, 9 (Spring):105-116 (2002).

Congress of the United States: Congressional Budget Office (CBO). Ed. Neil M. Singer. *Cost Growth in Weapon Systems: Recent Experience and Possible Remedies*. 11 October 1982. http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/51xx/doc5102/doc19-entire.pdf

Crumrine, Kevin T. *A Comparison of Earned Value Management and Earned Schedule as Schedule Predictor on DoD ACAT I Programs*. MS Thesis, AFIT-ENV-13-M-36. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, March 2013.

Crumrine, Kevin T. and Jonathan D. Ritschel. "A Comparison of Earned Value Management and Earned Schedule as Schedule Predictors on DoD ACAT I Programs." *Measurable News*, February 2013: 37-44 (2013).

Defense Acquisition University (DAU),*Defense Acquisition Guidebook*, 2013. https://acc.dau.mil/CommunityBrowser.aspx?id=488728

-----. *Gold Card*, August 2013. https://acc.dau.mil/CommunityBrowser.aspx?id=19577

Defense Contract Management Agency (DCMA). *DoD Earned Value Management Implementation Guide (EVMIG)*. Virginia: DCMA, 2006.

Department of Defense (DoD). *The Defense Acquisition System*. DOD Directive 5000.01. Washington D.C.: May 12, 2003.

Fleming, Quentin W. and Joel M. Koppelman. "Earned Value Project Management: A Powerful Tool for Software Projects," *Cross Talk, the Journal of Defense Software Engineering*, July 1998: 19-23 (1998).

-----. *Earned Value Project Management* (2nd Edition). Pennsylvania: Project Management Institute, 2000.

-----. "The Two Most Useful Earned Value Metrics: the CPI and the TCPI," *Cross Talk, the Journal of Defense Software Engineering*, December 2008: 16-18 (2008).

Government Accountability Office (GAO). *GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs*. Ed. Barry Leonard. DIANE Publishing, 2009.

Hayter, Anthony J. *Probability and Statistics for Engineers and Scientists* (3rd Edition). California: Duxbury, 2007.

Heise, Capt Scott R. *A Review of Cost Performance Index Stability*. MS thesis, AFIT/GCA/LSY/91S-12. School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB, OH, September 1991.

Henderson, Kym and Ofer Zwikael. "Does Project Performance Stability Exist? A Re-examination of CPI and Evaluation of SPI(t) Stability." *Cross Talk, the Journal of Defense Software Engineering*, Vol. 21 No. 4:7-13 (2008).

Hogg, Robert V. and Allen T. Craig. *Introduction to Mathematical Statistics* (5th Edition). New Jersey: Prentice-Hall Inc., 1995.

Kutner, M. H., Nachtsheim, C. J., & Neter, J. *Applied Linear Regression Models*. New York: McGraw-Hill, 2004.

Lipke, Walt. "Schedule is Different." *The Measurable News*, March & Summer 2003. http://www.earnedschedule.com/Docs/Schedule%20is%20Different.pdf

-----. "Re-Examination of Project Outcome Prediction:…using Earned Value Management Methods," *The Measureable News*, Summer:14-26 (2005).

McClave, James T., P. George Benson, and Terry Sincich. *Statistics for Business and Economics* (11th Edition). Boston: Prentice Hall Inc., 2011.

Office of the Under Secretary of Defense: Acquisition, Technology, and Logistics (USD-AT&L). *Office of the Secretary of Defense Central Repository for Earned Value Management Data*. Washington, D.C.: March 12, 2008.

Payne, Kirk I. *An Investigation of the Stability of the Cost Performance Index*. MS thesis, AFIT/GCA/LSY/90S-6. School of Systems and Logistics, Air Force Institute of Technology, Wright-Patterson AFB, OH, September 1990.

Project Management Institute- College of Performance Management (PMI-CPM).*Practice Standard for Earned Value Management*. Newton Square, PA: PMI, 2005.

Project Management Institute (PMI). *What is Project Management*, 2013. http://www.pmi.org/About-Us/About-Us-What-is-Project-Management.aspx

Popp, Michael. "Probability Distributions of CPI at Complete vs. CPI Today." Internal NAVAIR Report, Unpublished, 1996.

Ritschel, Jonathan D. "Efficacy of US Legislation in Military Acquisition Programs: Nunn-McCurdy Act Unveiled." *Economic Papers*, Vol. 31 No. 4:419-500 (2012).

Seibel, Michael. Personal Interview, 27 January 2014.

Smith, Greg J and Kym Henderson "Does Project Performance Stability Exist? A Re-examination of CPI and Examination of SPI(t) Stability," *Topical Workshop International Integrated Performance Management Conference,*14 (November 2008)
 http://www.evmlibrary.org/library/WS%2006%20-%20Program%20Performancwe%20Henderson%20SmithxPi%20Stability%20W orkshop%20Nov%202008%20v1%20FINAL.pdf

Society of Cost Estimating and Analysis (SCEA). "Module 15 Earned Value Management (EVM)," *Cost Estimating Body of Knowledge* (CEBoK), 2010.

Thickstun, Kristine and Edward White. "Predicting Over Target Baselines (OTBs)" *The Measurable News*, Issue 4:10-14 (2010).

Tracy, Steven P. and Edward D. White. "Estimating the Final Cost of a DoD Acquisition Contract," *Journal of Public Procurement* 11(2): 190-205 (2011).

Vita

Second Lieutenant Jacob L. Petter graduated from Reicher Catholic High School in Waco, Texas. He entered undergraduate studies at the United States Air Force Academy, where he graduated with a Bachelor of Science degree in Economics in May 2012 and was commissioned an officer in the U.S. Air Force. He was chosen to attend the Air Force Institute of Technology's Graduate School of Engineering and Management to earn a Master's of Science degree in Cost Analysis.

Upon graduation, he will be assigned to the Air Force Cost Analysis Agency at Joint Base Andrews in Maryland.

REPORT DOCUMENTATION PAGE

1. REPORT DATE *(DD-MM-YYYY)* 27-03-2014	**2. REPORT TYPE** Master's Thesis	**3. DATES COVERED** *(From – To)* Sep 2012 - Mar 2014
4. TITLE AND SUBTITLE An Analysis of Stability Properties in Earned Value Management's Cost Performance Index and Earned Schedule's Schedule Performance Index		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Petter, Jacob L, 2d Lt		**5d. PROJECT NUMBER** 11G222
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENV-14-M-49
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) College of Performance Management Attn: Kym Henderson 101 South Whiting Street, Suite 320 Alexandria, VA 22304 (703)370-7885 kym.henderson@gmail.com	**10. SPONSOR/MONITOR'S ACRONYM(S)** CPM **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The concept of a Cost Performance Index (CPI) "stability rule" originated with the seminal article from Christensen and Payne in 1992 and has become routinely cited by subsequent academic literature and EVM authors. A literature review reveals that the definition of what constitutes "stability" has morphed over time, with three separate definitions of "stability" permeating the literature. Additionally, while the original Christensen research found the cumulative CPI to be stable in 86% of DoD contracts (from 155 analyzed) at the 20 percent completion point, more recent research from Henderson and Zwikael (2008) questioned the generalizability of these findings. This research reexamines the question of CPI stability in a modern portfolio of DoD contracts utilizing all three definitions of stability. Next, this research examines potential stability in the Earned Schedule SPI(t) metric. The second stage of this research investigates whether there is a difference in CPI or SPI(t) stabilities between military services, contract types, acquisition life-cycle phases, or platforms. Comparison analysis executes tests on the median stability value for each category of contract. This research finds that CPI stability *both contradicts and supports* the stability rule depending on the stability definition used. The SPI(t) exhibits similar stability traits to CPI stability.

15. SUBJECT TERMS

EVM, Cost Performance, Schedule Performance, Stability

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Lt Col Jonathan D. Ritschel, AFIT/ENV
REPORT U	**ABSTRACT** U	**c. THIS PAGE** U	UU	107	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, x4441 jonathan.ritschel@afit.edu